Terminologie & Ontologie : Théories et Applications



Actes de la conférence

Articles acceptés pour publication sans présentation orale

TOTh 2011

Annecy – 26 & 27 mai 2011

Publications précédentes

TOTh 2007

Actes de la première conférence TOTh - Annecy - 1[er] juin 2007

TOTh 2008

Actes de la deuxième conférence TOTh - Annecy – 5 et 6 juin 2008

TOTh 2009

Actes de la troisième conférence TOTh - Annecy – 4 et 5 juin 2009

TOTh 2010

Actes de la quatrième conférence TOTh - Annecy – 3 et 4 juin 2010

Commandes à adresser à : toth@porphyre.org

Titre : TOTh 2011. *Actes de la cinquième conférence TOTh - Annecy – 26 & 27 mai 2011*
Editeur : Institut Porphyre, *Savoir et Connaissance*
http://www.porphyre.org
Annecy, 2011
ISBN 978-2-9536168-4-2
EAN 9782953616842

Terminologie & Ontologie : Théories et applications



Actes de la conférence

Articles acceptés pour publication sans présentation orale

TOTh 2011

Annecy – 26 & 27 mai 2011

avec le soutien de :

- Ministère de la Culture et de la Communication, Délégation Générale à la Langue Française et aux Langues de France
- Association Européenne de Terminologie
- Société française de terminologie
- Ecole d'ingénieurs Polytech'Savoie – Université de Savoie
- Association EGC (Extraction et Gestion des Connaissances)
- ISKO (International Society for Knowledge Organization) France



http://www.porphyre.org

Comité scientifique

Président du Comité Scientifique : Christophe Roche

Comité de pilotage

Rute Costa	Professeur, Universidade Nova de Lisboa
Loïc Depecker	Professeur, Université de Sorbonne nouvelle
Christophe Roche	Professeur, Université de Savoie
Philippe Thoiron	Professeur émérite, Université de Lyon 2

Comité de programme

Bruno Bachimont	Dir. Recherche, Univ. Technologie de Compiègne
Bruno de Bessé	Professeur, Université de Genève
Franco Bertaccini	Professeur, Université de Bologne
Gerhard Budin	Professeur, Université de Vienne
Teresa Cabré	Professeur, Universitat Pompeu Fabra, Barcelone
Marc van Campenhoudt	Professeur, Termisti, ISTI, Bruxelles
Danielle Candel	CNRS, Université Paris Diderot
Stéphane Chaudiron	Professeur, Université de Lille 3
Luc Damas	MCF, Université de Savoie
Sylvie Després	Professeur, Université Paris 13
François Gaudin	Professeur, Université de Rouen
Anne-Marie Gendron	Chancellerie fédérale suisse, Section terminologie
Jean-Yves Gresser	ancien Directeur à la Banque de France
Ollivier Haemmerlé	Professeur, Université de Toulouse
Michèle Hudon	Professeur, Université de Montréal
John Humbley	Professeur, Université Paris 7
Hendrik Kockaert	Professeur, Lessius Hogeschool (Anvers)
Michel Léonard	Professeur, Université de Genève
Pierre Lerat	Professeur honoraire, Equipe Condillac
Widad Mustafa	Professeur, Université de Lille 3
Fidelma Ní Ghallchobhair	Foras na Gaeilge (The Irish-Language Body)
Henrik Nilsson	Terminologicentrum TNC, Suède
Jean Quirion	Professeur, Université d'Ottawa
Renato Reinau	Suva, Lucerne
Gérard Sabah	CNRS, Orsay
Michel Simonet	CNRS, Grenoble
Marcus Spies	Professeur, Université de Munich
Dardo de Vecchi	Professeur associé, Euromed-Management

Comité d'organisation :

Responsable : Luc Damas
Samia Chouder, Joëlle Pellet

Avant propos



La Terminologie est un domaine scientifique par nature pluridisciplinaire. Elle puise, entre autres, à la linguistique, la théorie de la connaissance, la logique. Pour que cette diversité soit une richesse, il faut lui offrir un cadre approprié au sein duquel elle puisse s'exprimer et s'épanouir : c'est une des raisons d'être des Conférences TOTh.

Dans ce contexte, la formation et la transmission des connaissances jouent un rôle essentiel. La Formation TOTh, programmée sur un jour et demi précédant la conférence, se déroule depuis 2011 sur deux années consécutives dédiées pour l'une à la dimension linguistique et pour l'autre à la dimension conceptuelle de la terminologie, deux dimensions intimement liées.

La Disputatio, introduite à partir de cette année, renoue avec une forme d'enseignement et de recherche héritée de la scolastique. Elle vise, à travers une lecture commentée effectuée par un membre du comité scientifique, à donner accès à des textes jugés fondateurs de notre domaine, trop souvent oubliés voire ignorés.

La cinquième édition des Conférences TOTh a également été l'occasion de mettre en place un Prix « Jeune chercheur ». Décerné par le comité scientifique lors de la conférence, il récompense le travail soumis à TOTh d'un de nos jeunes collègues.

Notre collègue Michele Prandi, professeur à l'Università degli Studi di Genova, a ouvert la Conférence TOTh 2011 par un exposé passionnant sur : « Signes, signifiés, concepts : pour un tournant philosophique en linguistique ». Le ton était donné.

Ont suivi douze communications (hors conférence d'ouverture et disputatio) réparties sur deux jours en six sessions animées par différents présidents. Elles ont permis d'aborder en profondeur – chaque intervention dure au minimum 45 minutes – de nombreux sujets tant théoriques que pratiques rappelant qu'il ne peut y avoir de terminologie sans langue de spécialité ni savoir spécialisé.

Les douze communications, équitablement réparties sur les deux langues officielles de la conférence et provenant de sept pays différents, confirment l'audience internationale acquise aujourd'hui par TOTh.

Avant de vous souhaiter bonne lecture de ces actes, j'aimerais terminer en remerciant tous les participants de TOTh 2011 pour la richesse des débats et des moments partagés.

Christophe Roche

Président du comité scientifique

Table des matières

Fouille de l’oral tel qu’il est parlé

Kafil Hajlaoui*, Rahma Boujelbane*, Fadila Bentayeb*,
Carole Etienne**, Omar Boussaid*

*Université de Lyon (ERIC Lyon 2)
5 Avenue Pierre Mendès-France, 69676 Lyon - France
{prenom.nom}@univ-lyon2.fr
**Ens de Lyon
15 parvis René Descartes - BP 7000 69342 Lyon - France
carole.etienne@ens-lyon.fr

Résumé. L’analyse des corpus de l’oral transcrit note une série de phénomènes qui perturbe le traitement automatique. Cet article présente une contribution qui vise à identifier automatiquement des phénomènes complexes qui composent une interaction (remerciements, conflits, plaisanteries familières, ...). L’objectif de ce travail est de proposer de nouveaux instruments permettant d’enrichir les outils d’analyse et d’interrogation en ligne de CLAPI (Corpus de Langue parlée en Interaction). Les difficultés de l'analyse de l’oral afin d'en extraire les séquences nous ont conduit à mettre en œuvre et à nous approprier de nombreuses techniques informatiques très diverses: fouille de données, traitement de la langue naturelle etc. Ces techniques ont été intégrées au sein d'une architecture de traitement originale qui a permis de détecter les séquences argumentatives pour l’oral transcrit. Ces résultats ont été évalués et validés par un expert linguistique du domaine.

1. Introduction

Les corpus de données authentiques, enregistrées dans leurs sites naturels de production, constituent aujourd'hui une ressource fondamentale pour de nombreuses disciplines en sciences humaines et sociales.

Cette ressource a donc acquis au fil de la dernière décennie une valeur à la fois scientifique, historique (importance de stocker ces traces labiles de pratiques dans des archives appropriées) et patrimoniale. La notion de patrimoine incluant aussi la documentation de la vie quotidienne et non seulement la préservation de traces monumentales.

Le laboratoire ICAR[1] est reconnu internationalement dans ce domaine en mettant à disposition de la communauté scientifique une des rares bases de données de corpus authentiques enregistrés en audio ou en vidéo : CLAPI2. Les corpus constituent des objets complexes aussi bien pour les données primaires (enregistrements multisources, documents manipulés au cours de l'interaction) que secondaires (transcriptions avec différents niveaux de granularité, métadonnées pour expliciter le contexte de l'interaction). Durant l'année 2004, un prototype développé en collaboration avec le laboratoire ERIC pour modéliser ces données et assurer leur mise en ligne en respectant les différents niveaux d'accès [1]. Dans une seconde phase, CLAPI a mis à disposition un ensemble d'outils de requêtes, intégrant les caractéristiques de la parole et de l'interaction comme la temporalité ou la multimodalité, pour assister le chercheur dans les différentes phases de son analyse.

Pour poursuivre cette approche exploratoire, les informaticiens et les linguistes cherchent à concevoir de nouveaux instruments capables de traiter de telles données pour identifier de nouveaux objets interactionnels de manière (semi)automatique. À partir de cette base de données complexes CLAPI et des problématiques des chercheurs, un premier projet interne à l'Université Lyon 2 en collaboration avec le laboratoire ICAR a été défini en recherche d'informations à partir de données complexes. Il porte sur la détection automatique de séquences, détection que les outils actuels ne savent pas effectuer.

Les transcriptions du français parlé notent une série de phénomènes qui les distinguent du français écrit standard. Ces phénomènes, qualifiés traditionnellement de disfluences (répétitions, euh, etc.), perturbent la linéarité de l'énoncé et constituent un problème pour le traitement automatique des langues [2]. Puisque l'annotation devient difficilement automatisable.

[1] ICAR (Interaction, Corpus, Apprentissage, Représentation)
[2] http://clapi.univ-lyon2.fr/

Vis-à-vis de notre objectif d'extraction d'information pertinente dans le corpus, cette dernière se caractérise par un contenu informationnel étroitement lié au thème traité (problème d'hétérogénéité des thèmes considérés dans le discours). Ce contenu sera analysé par le biais de la détection de séquences au cours d'une interaction (thème, conflit,...). *Une séquence est une action sémantique complète autour d'un matériel verbal (verbe, phrase, texte) : séquence émotionnelle, séquence argumentative, question-réponse, proposition-acceptation*... [3]

Pour tenir compte de ces contraintes, notre approche commence donc par un traitement linguistique des disfluences sur le plan morphologique et syntaxique sans toutefois dénaturaliser les spécificités de l'oral. Après ce pré-traitement du corpus, nous proposons une approche permettant le repérage des séquences et leur annotation sémantique. Cette démarche est basée sur des méthodes de *text mining* et de traitement automatique de la langue.

Dans la suite de cet article, nous présentons la base de données CLAPI et les différentes conventions d'annotation qui composent les phénomènes de l'oral. Dans la troisième section, nous présentons un état de l'art sur des travaux qui ont traité la fouille de l'oral avec ses différents disfluences pour positionner ensuite notre approche hybride mêlant des méthodes de *Text ming* et du Traitement Automatique de Langue (TAL). Nous finalisons avec nos conclusions de ce travail et les différentes perspectives qui sont ouvertes suite aux résultats prometteurs obtenus.

2. La base de données CLAPI

La plateforme CLAPI du laboratoire ICAR est un environnement d'archivage et d'analyse de corpus d'interactions enregistrées en situation authentique. Les corpus hébergés dans CLAPI ont été recueillis dans des contextes de conversations ordinaires, mais également dans différents types d'institutions, de services publics ou d'entreprises privées (poste, mairies, études notariales, commerces, etc.), dans l'enseignement secondaire et universitaire, ou dans différents contextes médicaux (médecin, dentiste).

Dans son état actuel, la plateforme CLAPI comporte environ 300 enregistrements soit 170h de données regroupées en une cinquantaine de corpus, ainsi qu'un ensemble d'outils d'exploitation pour leur analyse qualitative et quantitative. Le site Corinte[3], dédié à la recherche sur les corpus de langue parlée en interaction est associé à cette initiative pour expliciter la démarche qui accompagne la constitution d'un corpus, du recueil à la mise en ligne.

Le traitement des corpus réunis dans la plateforme CLAPI présuppose un certain nombre de formats standard, notamment en ce qui concerne la caractérisation des

[3] http://icar.univ-lyon2.fr/projets/corinte/

corpus. CLAPI comporte 75 entrées hiérarchisées (génériques ou spécifiques, obligatoires ou facultatives) couvrant les points suivants [1]:

Métadonnées	Modèle
Informations générales	Nom du corpus, texte de présentation, dates et lieux de recueil, liste des corpus associés, durée totale…
Auteurs	Responsable, collecteurs, transcripteurs…
Genre interactionnel	Interaction privée, de travail, médicale…
Enregistrements	Nom, date et lieu de recueil, durée, type de support (audio vidéo), anonymisation…
Transcriptions	Convention et logiciels utilisés, orthographe standard ou adaptée, alignement et balisage, exhaustivité, anonymisation…
Locuteurs	Nom ou pseudonyme et caractérisation sociolinguistique

TAB. 1 – *Métadonnées de la base CLAPI*

Au-delà des métadonnées, le contenu des transcriptions a été modélisé en XML. Cette modélisation est faite à partir de la convention fournie par le transcripteur et ce quelque soit la convention de transcription utilisée. Elle tient compte également de la vérification manuelle de la médiathèque d'ICAR pour les notations problématiques (un même signe pour noter différents phénomènes, des oublis, plusieurs notations pour le même phénomènes, …), et d'un traitement informatique itératif composé d'un parseur, d'une chaîne de validation du lexique et des phénomènes interactionnels les plus courants avant leur insertion dans la base.

Les catégories de notation actuellement prises en compte au niveau de la totalité des corpus CLAPI sont les phénomènes de base du discours oral :

production verbale/tour de parole;

chevauchement ;

pause/silence ;

token/mot ;

Pour annoter les transcriptions, le laboratoire ICAR a mis au point et documenté sa propre convention de transcription ICOR[4].

Un nouveau potentiel de recherche s'impose pour contribuer au traitement de ce type de données complexes. Les chercheurs de différentes disciplines commencent à s'intéresser au traitement automatique de ce type de données.

[4] http://icar.univ-lyon2.fr/documents/ICAR_Conventions_ICOR_2007.pdf

3. Positionnement et présentation de l'approche

3.1 Etat de l'art

L'étiquetage morpho-syntaxique d'un texte est une étape fondamentale de son analyse et préliminaire à tout traitement de plus haut niveau. Des étiqueteurs fiables existent pour le français. Cependant, ils sont conçus pour les textes écrits, et sont de ce fait mal adaptés aux spécificités d'une langue moins normalisée [4].

Pour étiqueter l'oral, des travaux antérieurs assez récents ont été réalisés mettant en exergue des méthodologies ou des approches pour pallier les difficultés causées par certaines disfluences et automatiser le traitement de l'oral. Pour le corpus Media (corpus de dialogue en langue français) [5], les auteurs ont proposé un processus d'interprétation automatique pour composer les structures sémantiques à partir des composantes sémantiques de base, utilisant des modèles contenant des unités lexicales et conceptuelles.

D'autres travaux se sont focalisés sur le corpus ESLO [6], constitué par la transcription d'enregistrements oraux, et présente des particularités mal prises en compte par les étiqueteurs standard. Cette approche repose sur la construction d'un étiqueteur par apprentissage automatique, à partir d'un corpus de référence étiqueté par Cordial (Logiciel de correction de l'orthographe et de la grammaire, Correcteur d'orthographe et correcteur de grammaire) et corrigé manuellement. D'autres travaux portent sur le processus d'indexation et d'annotation automatique de grands flux de paroles issus d'émissions télévisées ou radiophoniques et décrivent les principaux problèmes qui se posent à l'analyse : traitement des disfluences de l'oral spontané, mais également la gestion des erreurs dues aux étapes antérieures de reconnaissance de la parole et d'étiquetage morphosyntaxique [7]. Une autre étude s'attaque à l'étape suivante de la chaîne de traitement, c'est-à-dire l'étiquetage morphosyntaxique de corpus déjà transcrits [8]. Cette opération est délicate, car les systèmes existants, généralement développés et entraînés pour l'écrit se heurtent aux particularités des transcriptions de l'oral, telles que l'absence de ponctuation, ou la présence de répétitions et de divers phénomènes de production. Par exemple, la répétition avec correction du genre du déterminant « le la » risque d'être étiquetée DET + NOM (comme le la du diapason).

L'analyse du parlé spontané a pris aussi un tournant important avec la méthode de Blac et ses collègues [9]. Son approche est basée en premier lieu sur une étape de prétraitement qui consiste à reformater et à marquer des énoncés qui brisent la structure syntaxique du texte, et en second temps sur les modules de normalisation et de chunking. Cordial offre la possibilité de détecter les mots répétés (ce qui est une de ses fonctionnalités en tant que correcteur orthographique et grammatical). Les répétitions immédiates concernant un mot unique tel que « pour pour » sont en général bien détectées et la seconde instance est supprimée, ce qui conduit à une étiquette correcte sur la première [10].

3.2 Objectifs

« *On ne parle pas de la même façon à un ami dans un café qu'à son directeur dans une salle de réunion* ». Les relations et les statuts sociaux sont différents entre les personnes. Parler c'est adopter une conduite langagière qui correspond à une situation de la communication. Dans son fond, cette adaptation est différente d'une situation à une autre. Mais sur la forme, les locuteurs utilisent toujours les mêmes traits de communication pour se comprendre : un « oui oui » pour exprimer son accord, un « non non » pour exprimer son conflit, un « euh » pour l'hésitation … L'oral transcrit est un langage où la terminologie est très spécifique et les frontières entre les mots sont souvent peu évidentes. La réalisation d'un phénomène linguistique dans une production verbale dépond fortement de son contexte. On estime bien que l'étude sur le choix de la terminologie employée dans le discours joue un rôle important dans la détection et l'évaluation de la pertinence de l'information portée par le contexte

Ce format des transcriptions et leurs particularités rendent difficiles les traitements automatiques de ce type de données. Au contraire des méthodes de traitement de l'oral, qui s'appuient sur des démarches nettement statistiques ou linguistiques, notre étude a comme objectif de combiner les deux méthodes. Une première méthode statistique basée sur le comptage des occurrences à la fois utilisée pour le prétraitement (la correction des disfluences) ainsi la détection des thématiques abordées dans les productions verbales. La deuxième méthode (TAL) basée sur l'utilisation des ressources linguistiques tel que les connecteurs et les conjonctions pour analyser les transcriptions afin d'extraire un bloc d'information pertinentes appelées « séquence ». Ainsi, notre étude montrera que la conversation est une communication entre trois niveaux : les mots de l'oral sont corrélés avec la suite des mots de la production verbale pour définir une séquence. L'ensemble de séquence utilisé dans la production verbale fait émerger un type de conversation (thème de conversation). Cette communication établie se base sur l'hypothèse que pour une tâche de discussion de meilleure qualité, le locuteur cherche à repérer les marques d'affirmation, de certitude, de modalisation, de doute, etc, et les mettre en parallèle pour discuter des phénomènes qu'elles révèlent. La performance de communication est fortement liée à la capacité de mise en œuvre effective de la compétence linguistique du locuteur à chaque intervention.

4. Fouille de l'oral

4.1 Pré-traitements des disfluences

Le prétraitement des textes est utilisé de façon à améliorer la qualité du texte à traiter sans modifier l'information qu'il contient. Ce qui permet d'affiner les résultats de l'application des méthodes informatiques.

Autocorrection automatique

Nous appelons autocorrection (immédiate) le phénomène langagier qui consiste, pour un locuteur, à énoncer un morphème suite à un autre morphème différent qui appartient à la même catégorie grammaticale. Ce deuxième morphème vise à corriger le premier morphème énoncé. Voici un exemple type d'autocorrection : *Le La migration.*

La terminologie est loin d'être unifiée. Et l'on rencontre dans la littérature le terme d'autocorrection pour n'importe quelle séquence dans laquelle le locuteur se corrige, indépendamment de la forme que peut prendre cette correction. [11] utilisent quant à eux le terme auto-réparation.

En ce qui nous concerne, nous avons opté en premier lieu à une analyse morpho-syntaxique via l'usage de l'analyseur syntaxique *tree-tagger 5.* C'est un outil qui permet d'annoter un texte avec des informations sur les parties du discours genre de mots: noms, verbes, infinitifs et particules et des informations de lemmatisation. Si on trouve une suite de natrure gramitacale suivante : det+det+nom, on élémine le premier et on garde le deuxième. En l'occurrence, « le la migration » devient « la migration ».

La correction des répétitions dans l'oral peut être aussi porteuse d'information pertinente dans le discours entre les deux locuteurs : Exemple : « *oui oui oui* »

Dans ces cas nous n'avons pas besoin de corriger cette répétition parce qu'elle montre bien que les deux locuteurs sont en accord sur le point de discussion. Dans d'autre cas, comme « *non non non* », elle peut au contraire montrer une situation de conflit entre les locuteurs. Le phénomène de répétition n'est pas toujours soumis à correction, il peut être une des spécificités de l'oral qu'il faut au contraire garder dans le corpus [2].

Trucage orthographique (élision « ` »)

L'élision est un phénomène de l'oral : c'est le fait de ne pas prononcer une voyelle lorsqu'elle précède une autre voyelle ou d'un « *h* » muet par exemple (*d'autre* non *de autre*, *aujourd'hui* non *aujourde hui*). En français, l'élision peut être marquée à

[5] http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html

l'écrit par une apostrophe. En fouillant nos corpus, nous observons différents types d'élision qui ne se réfèrent pas à cette définition et se placent dans n'importe qu'elle position dans le mot. Cette "élision non standard" se traduit par une déformation orthographique et affecte aussi bien les pronoms, les suites « pronom+verbe » que les lexèmes exemples : « *`fin* » au lieu de « *enfin* », « *not`* » au lieu de « *notre* », « *p`t-être* » au lieu de « *peut être* ».

Repérer la marque de l'élision *(`)*.

Rechercher dans la dictionnaire de CLAPI (i) le mot le plus semblable en terme de caractères, en chercher le pourcentage le plus élevé de ressemblance entre deux mots ; (ii) choisir le mot le plus fréquent.

Les chevauchements de parole

Dans les conversations, il est très fréquent que deux (ou plusieurs) locuteurs parlent en même temps. On a alors des séquences où la parole est superposée. Dans des études interactionnelles, il est important de garder une trace de ces chevauchements, et pas uniquement des paroles prononcées. En effet, quand le chevauchement est interne au tour de parole (c'est-à-dire qu'un locuteur B commence à parler alors qu'un autre locuteur A est déjà en train de parler, mais sans prendre définitivement la parole puisque le locuteur A poursuit son propos). Nous avons adopté l'idée de [9] pour faire la correction en réorganisant les tours de parole des locuteurs dans un bloc de chevauchement. De cette manière, on garde la sémantique d'une phrase permettant l'analyse de chevauchement qui représente un phénomène important dans le corpus oral porteur d'informations.

4.2 Détection de séquence

Une séquence peut être définie comme une action sémantique complète autour d'un matériel verbal (mot, phrase, texte). C'est le volume du texte qui permet de décrire et de délimiter une tranche de discours autour d'une suite émotionnelle, argumentative, Q-R, proposition-acceptation, un acte de langage ou un thème.

En se basant sur l'expertise linguistique, nous avons pu identifier trois types de séquence :

1. *Séquence thématique* : c'est une séquence qui permet de décrire le thème abordé dans le discours. Une production verbale peut être réalisée autour de plusieurs thèmes.
2. *Séquence complexe* : c'est une séquence qui permet de décrire les phénomènes linguistiques complexes dans le discours (argumentation, suite émotionnelle, conflit…)
3. *Séquence simple* : ce sont les mots de l'oral (les interjections, les connecteurs…) qui peuvent déclencher une séquence complexe ou un phénomène linguistique simple.

La détection des séquences est un problème délicat à traiter en l'absence d'une grammaire pour l'oral [12,13,14]. Les transcriptions de notre corpus doivent être soumises à un traitement sophistiqué qui permet de rapprocher l'oral transcrit d'un texte écrit tout en gardant les spécificités de l'oral.

Nous avons commencé par une phase d'étude de la définition de la séquence et la manière de la repérer dans le texte. L'objectif était de mettre un protocole expérimental pour définir une séquence à partir de l'analyse du comportement de l'expert et de comprendre les fondements de cette détection. Les questions auxquelles on a tenté de répondre dans notre étude sont :

- Comment l'expert sélectionne les termes de références décrivant la notion de séquence ?
- Comment l'expert délimite la production verbale d'une séquence ?
- Comment l'expert interprète les séquences déjà détectées avec de nouvelles séquences?

Grâce à cette expérience manuelle, nous essayerons ensuite d'orienter, d'adapter ou de reconstruire des outils et des méthodes en linguistique ou en *Text mining* pour reproduire le raisonnement de l'expert

Notre travail est guidé par des besoins et des hypothèses de l'oral. La mise en place des solutions de détection des séquences est construite dans un but pragmatique décrit principalement par la définition de la séquence et les types d'informations nécessaires pour la déterminer. Rappelons toujours que notre objectif est d'enrichir le corpus de l'oral transcrit par des nouvelles balises sémantiques décrivant des séquences (phénomènes linguistiques). Les problèmes posés par les caractéristiques de ce corpus sont nombreux, que ce soit au niveau du pré-traitement ou de l'interrogation. Face à ces problèmes, nous avons étudié des solutions spécifiques. Dans une première étape, nous commençons par un traitement sur le texte transcrit (voir section 2). Cette étape a comme objectif d'adapter le texte de l'oral à un texte bien écrit en corrigeant les disfluences (élision, auto-correction…) sans dénaturer l'oral et ses spécificités.

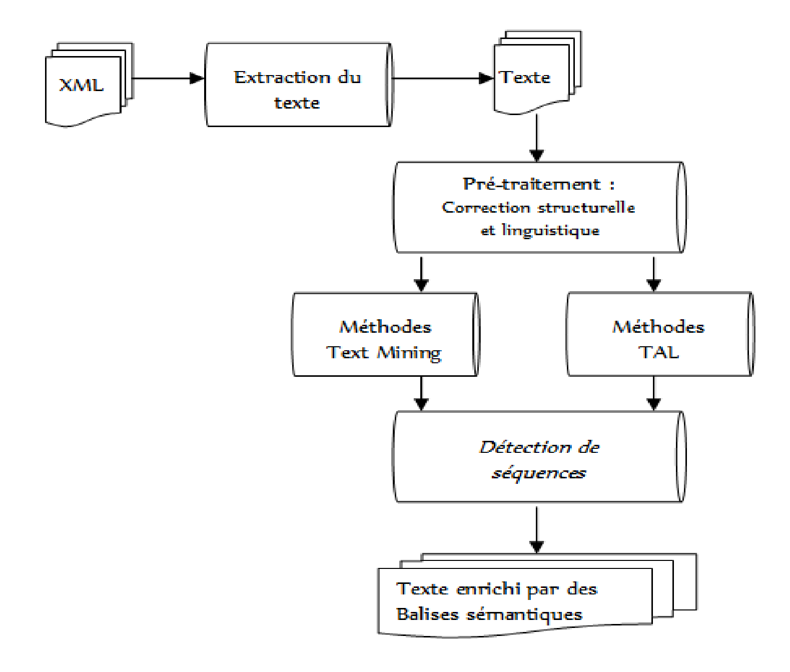


FIG. 1 – *Processus de détection de séquence.*

La deuxième étape consiste à appliquer les méthodes d'analyse de *Text mining* et de TAL. L'algorithme OKM (*Overlapping K-Means*) [11] est appliqué pour extraire les thématiques détectées dans chaque production verbale (section détection des topics). A chaque production verbale est associé un thème dominant. La méthode de TAL est appliquée pour repérer et extraire les séquences simples et complexes. Le résultat de ces deux méthodes est un ensemble de nouvelles annotations sémantiques qui vont être ajoutées au corpus XML initial. A la fin nous obtenons une nouvelle version annotée sémantiquement. Il est nécessaire de préciser que notre approche garde toujours les spécificités de l'oral qui sont traduites par des balises dans le corpus initial ou par d'autres phénomènes traduits par des conventions de transcriptions (pauses, chevauchements …)

4.3 Les patrons syntaxiques

Les patrons linguistiques sont le résultat de la construction d'une signature contextuelle. L'utilisation de cette technique est basée sur les principes de la sémantique distributive qui admet que la signification d'un mot est fortement corrélée aux contextes dans lesquels il apparait. D'une façon plus élaborée, un patron lexico-

syntaxique identifie la relation recherchée plus précisément en définissant également des contraintes syntaxiques ou typographiques sur le contexte des termes [15].

Les patrons linguistiques consistent à schématiser le contexte lexical et syntaxique commun à certains fragments de texte. Cette schématisation constitue le patron lexico-syntaxique en lui-même, qui pourra être utilisé pour extraire des fragments de texte à partir d'un corpus textuel [16].En linguistique, les approches par patrons sont utilisées pour associer des régularités structurelles à des informations sémantiques. C'est Hearst qui a été la première à utiliser cette méthode dans le contexte de l'extraction d'information. Elle a proposé des ensembles de patrons lexico-syntaxiques qui sont facilement repérables dans un texte et qui apparaissent fréquemment dans le but de reconnaitre certaines relations lexicales sans ambigüités. Hearst montre à partir de l'exemple de la phrase : « *The bow lute, such as the Bambara ndang, is plucked […]* » sans savoir ce que sont un *Bambara ndang* et un *bow lute*, le lecteur est capable d'indiquer qu'un *Bambara ndang* est une sorte de *bow lute*. Dans cette phrase la relation d'hyponymie peut être reconnue par le patron suivant : « un terme suivi par '*such as*' et un autre terme ». Elle est abstraite au sein du patron suivant :

X such as Y, où X et Y sont des syntagmes nominaux

Dans notre cas, les patrons généralisent des expressions lexico-syntaxiques identifiées dans les textes et fournissent ainsi une formalisation générique de relations lexicales particulières identifiées dans le corpus. Ils sont créés par l'abstraction des différentes instances apparaissant au sein du corpus, pour chacune de ces relations lexicales.

L'approche de détection automatique des séquences complexes se base sur la génération et la projection des patrons syntaxiques utilisant le système de traitement automatique de la langue naturelle UNITEX[6]. La sortie est un corpus annoté sur la base des occurrences des patrons retrouvés dans le texte.

[6] http ://www-igm.univ-mlv.fr/˜unitex/

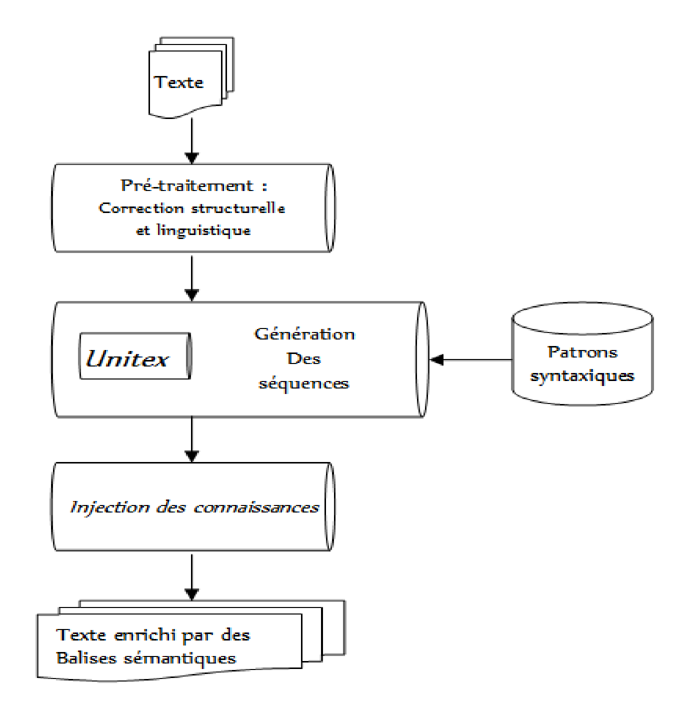


FIG. 2 – *Approche de détection des séquences complexes.*

Unitex offre un cadre de travail très intéressant et accessible au linguiste non informaticien. Il permet la formalisation graphique des automates sous la forme de grammaires locales. Il offre un nombre considérable d'outils permettant leur application au texte. Pour une description complète de ces outils, nous renvoyons le lecteur au manuel d'Unitex [17].

5. Expérimentation et résultats : application au cas de CLAPI

5.1 Détection des Topics

Les méthodes de classification actuelles ne sont pas adaptées à la recherche d'une organisation des données en classes avec recouvrement. Ce type de schéma de classification devient pourtant indispensable pour appréhender les domaines d'applica-

tion actuels tels que nos transcriptions des corpus oraux. Cette proposition s'appuie d'une part sur la définition d'un critère objectif permettant d'évaluer les recouvrements, et d'autre part sur une méthode d'exploration de cet espace des possibilités. Cette méthode a été mise en œuvre par l'application de l'algorithme OKM (*Overlapping K-means)*. Cette application a donné des résultats convaincants après un prétraitement cohérent du corpus. En effet OKM [18] a permis la détection de thématiques autour d'une production verbale. Nous rappelons que le choix des différentes mesures (*tf, tf*idf, p/a*) a une influence significative sur les performances de nos résultats.

Corpus d'application : "Négociation sur les loyers – Commission de conciliation : Amélioration des abords". Ce corpus concerne une séance de conciliation dans le cadre d'un litige sur l'augmentation du prix d'un loyer. Les locuteurs engagés dans cette interaction sont : le locataire, le propriétaire, les membres d'une commission de conciliation. Le tableau 1.2 montre le résultat de l'application de l'algorithme OKM sur le corpus "Amélioration des abords ".

Mesure	classes
tf	0=Monsieur propose, 1=décret, 2= monsieur, 3=veut, 4=question, 5=avis, 6= charges, 7= 8=sais, 9= attendez, 10= ben, 11= plantation, 12= bail, 13= heu, 14= francs, 15= écoutez.
*tf*idf*	0=Bas, 1= charge, 2= sûr, 3= bon ben, 4= équipement, 5= deu, 6= octobre, 7= ent, 8= bail, 9=heu, 10= écoutez, 11=augmentation, 12= sommes, 13=choses,1 4=monsieur.
p/a	0=Mh, 1= di frans, 2= conciliation, 3= heu, 4=bon ben, 5= madame, 6= monsieur, 7= écoutez, 8=question, 9= bail, 10= décret, 11= monsieur le président, 12= parce que, 13= abord, 14= loi quillot.

TAB. 2 –*Résultat de la détection des topics avec OKM*

Pour l'évaluation de ces résultats, nous avons eu recours à un expert linguistique pour juger la pertinence des classes obtenues. En effet, dans un premier constat, nous avons remarqué qu'il y a une possibilité d'éliminer les petits mots de l'oral (*heu, ben….*). Ces catégories de mots ne contribuent pas réellement à la formation du sens des thèmes pertinents dans les productions verbales.

Par une interprétation plus approfondie de l'expert, il s'est avéré que les verbes modaux *valoir ,savoir…* peuvent être des thèmes pertinents. En effet le verbe modal indique une volonté de passer d'un monde actuel (avec ses caractéristiques) à un autre monde possible (avec d'autre caractéristiques). Le traitement de notre corpus a pu extraire par exemple le thème "*savoir vouloir*", on peut alors dire que le monde

actuel serait "les loyers sans augmentation" et le monde possible serait "les loyers avec augmentation". Entre ces deux mondes, plusieurs phénomènes linguistiques, cognitifs de l'oral existent et ils sont traduits par différentes formes. C'est pourquoi dans l'extraction des classes thématiques, OKM fait apparaître ces thèmes.

Pour l'ensemble des autres thèmes, on a pu dégager deux catégories : une première basée sur le fond de la discussion, les thèmes qui sont abordés, les objets de la discussion. Comme par exemple "*On parle du décret, du bail ou de charges etc…*". Une deuxième catégorie qui est plutôt basée sur la forme de la discussion, le genre et les caractérisations de la discussion, la manière dont on a abordé le thème de fond comme "*madame, monsieur propose*". Il s'agit de caractériser et de spécifier le genre de la discussion. Ce thème peut nous donner de l'information sur le déroulement de l'interaction.

Ce raisonnement reflète bien deux aspects de l'interaction de l'oral traduisant des spécificités qui concernent à la fois le fond et la forme de la discussion.

5.2 Détection de séquences argumentatives

La génération de variantes de patrons a pour rôle d'étendre la couverture du système en proposant des structures sémantiquement équivalentes. Cette étape se base sur l'expertise humaine ainsi que sur le résultat de l'annotation des mots de paroles [19]. Autour de chaque mot pertinent, nous proposons un schéma syntaxique traduit par un patron qui permet d'encadrer l'information pertinente.

De nombreuses études ont mis en évidence l'adéquation des automates aux problèmes linguistiques [20, 21, 22]. Ainsi, une grammaire décrit des séquences de mots et produit des informations linguistiques (sur la structure syntaxique par exemple). Un dictionnaire représente les séquences de lettres et produit les informations lexicales associées. Le transducteur d'un texte représente les séquences de mots qui composent chaque phrase et leur associe des informations lexicales ou syntaxiques des résultats produits par différentes analyses. Les grammaires sont représentées au moyen de graphes que l'utilisateur peut créer et mettre à jour. L'application de dictionnaires à un texte consiste à construire l'union des transducteurs de chaque dictionnaire avec le transducteur du texte. Une grammaire locale est une représentation par automate de structures linguistiques difficilement formalisables dans des tables de lexique-grammaire ou dans des dictionnaires électroniques. Les grammaires locales, représentées sous la forme de graphes, décrivent des éléments qui relèvent d'un même domaine syntaxique ou sémantique.

Les descriptions linguistiques, décrites sous la forme de grammaires locales, sont utilisées pour une grande variété de traitements automatiques appliqués sur les corpus de texte. Ces grammaires locales sont un moyen puissant de représenter la plupart des phénomènes linguistiques. Ce sont des variantes des grammaires algébriques, également appelées grammaires hors-contexte. Une des principales fonctionnalités d'Unitex est la recherche d'expressions dans des textes. Une fois que

le texte a subi une opération de prétraitement (normalisation des formes non ambigües, découpage de texte en phrases) et que les dictionnaires électroniques ont été appliqués, on peut effectuer des recherches sur ces textes en leur appliquant les grammaires.

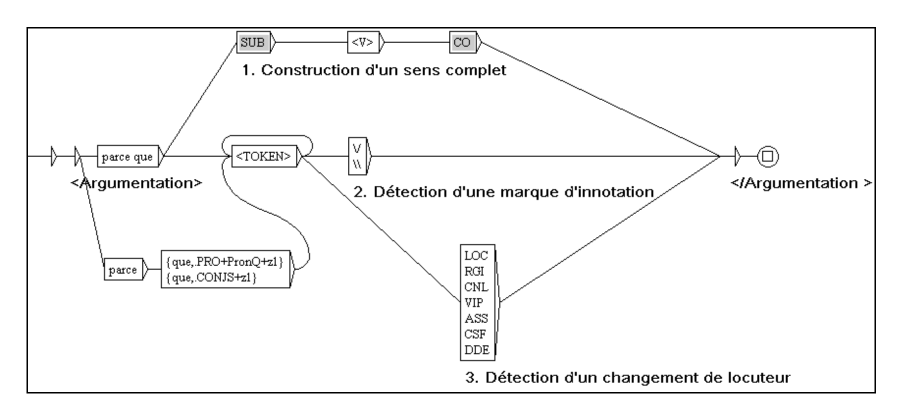


FIG. 3 – *Patron syntaxique transcrit sous Unitex pour détecter une séquence d'argumentation*

La détection de la séquence argumentative est analysée suivant les trois cas :

On cherche à construire un sens complet de la phrase <Sujet> <verbe> <Complément>. Exemple si on a la séquence argumentative :

"*parce que c'est en quatre-vingt-deux que ça aurait du s'appliquer.*"

On cherche à finir l'argumentation par une marque d'intonation qui dans la plupart des cas marque la fin de la justification.

On cherche à détecter un changement de locuteur. Dans le cadre de notre transcription, nous avons les pseudos suivants : LOC, RGI, VNL, VIP, ASS, CSF, DDE.

La projection des patrons sur le corpus se fait par la recherche des occurrences des schémas linguistiques, traduite sous la forme d'automates, dans le texte. On se base sur le module *locate* d'Unitex qui permet cette projection.

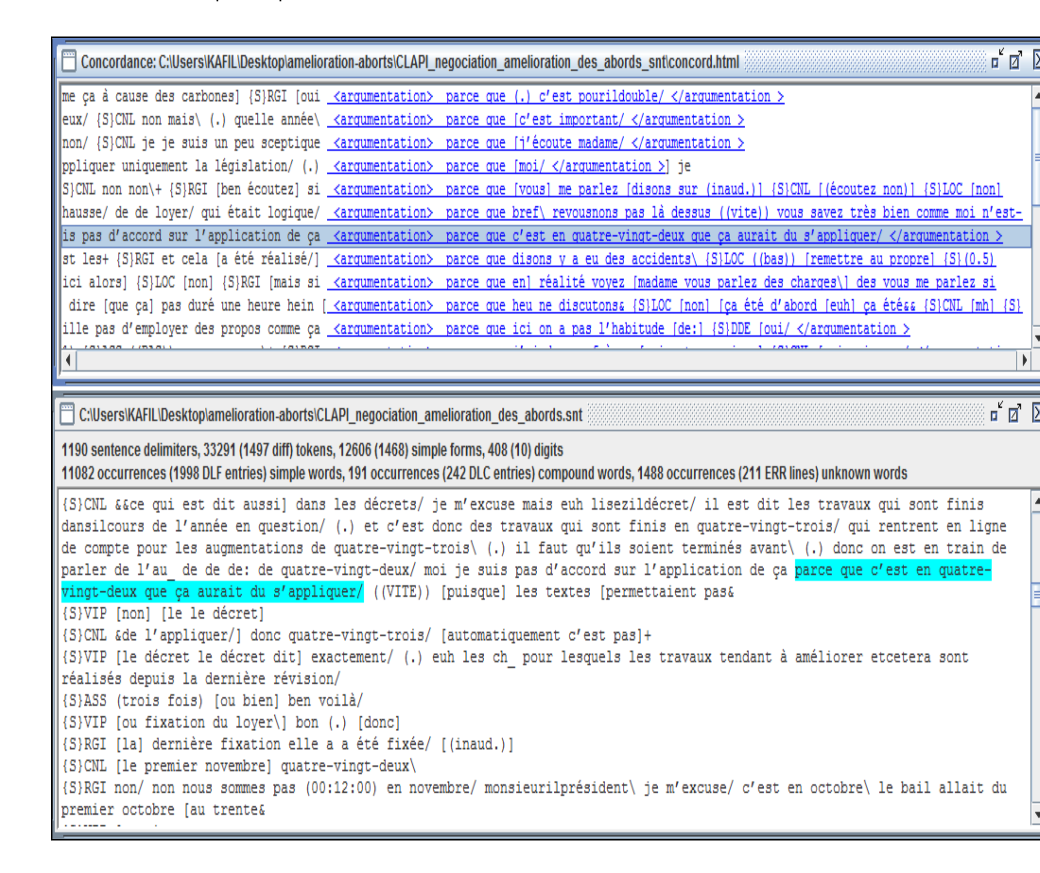


FIG. 4 – *Exemple de projection de patrons pour détecter des séquences d'argumentation*

La figure ci-dessus montre bien que tout les contextes d'argumentation exprimés par le connecteur « parce que » sont localisés. Chaque occurrence détectée est délimitée entre la balise <argumentation> ….</argumentation>

6. Discussion

La recherche et l'extraction d'informations pertinentes, à partir de données fournies en formats hétérogènes et relevant du domaine public ou privé, requiert l'utilisation de mécanismes avancés permettant de manipuler la syntaxe mais également la sémantique des informations. Dans cet objectif, les travaux présentés dans ce papier ont visé à confronter les techniques de la recherche et l'extraction d'information à des données en sciences humaines. Le point de départ de notre travail était un pro-

blème formulé dans un contexte de sciences humaines, qui porte sur le traitement automatique des transcriptions de l'oral pour la génération des connaissances. Tout au long de ce travail, nous avons présenté un ensemble d'observations, d'hypothèses, de réalisations et d'évaluations. En effet, nous avons proposé une approche descriptive complète qui va au-delà d'un simple mot pour effectuer des analyses sémantiques afin de faire émerger des phénomènes linguistiques complexes. Trois niveaux d'annotations sémantiques sont étudiés : annotation thématique pour détecter le thème de chaque production verbale, annotation des phénomènes complexes pour détecter les séquences dans les productions verbales (séquence argumentative, séquence de conflit…) porteuses d'une information pertinente, annotation des mots de paroles (les interjections, les marqueurs de relation…)

L'annotation s'avère difficilement automatisable compte tenu des disfluences rencontrées fréquemment dans le corpus. Notre contribution a commencé alors par un traitement de ces phénomènes linguistiques sur le plan morphologique et syntaxique sans toutefois dénaturer les spécificités de l'oral. Après cette étape de pré-traitement du corpus, nous continuons par une approche permettant le repérage des séquences et l'annotation sémantique. Des résultats expérimentaux ont été obtenus pour chacune de ces étapes et ont été évalués par une expertise humaine.

Les résultats présentés nous encouragent à poursuivre nos investigations avec un jeu de données plus important, dans des situations interactionnelles différentes pour valider la méthodologie et entériner les outils par un passage à l'échelle. Concernant les outils, de nouvelles adaptations seront proposées pour inclure d'autres spécificités de l'oral plus complexes comme la temporalité ou la multimodalité. En parallèle, on cherchera à enrichir la recherche des thèmes à partir des observables des linguistes et à contribuer ainsi aux travaux menés en sémantique de l'oral.

7. Conclusion et perspectives

Nous avons présenté dans cet article une méthodologie de détection des phénomènes linguistiques à partir des corpus de textes de transcriptions d'oral. Des expérimentations ont été présentées et qui ont montré l'intérêt de tirer profil de la terminologie caractérisant l'oral pour extraire l'information pertinente. Cette méthodologie inclura une étape de validation des séquences détectées automatiquement par les outils pour transformer le résultat "séquence" en une donnée fiable, intégrée à la transcription, disponible pour les outils de requêtes au même titre que les phénomènes de base comme les chevauchements ou les pauses. Ainsi, les linguistes auront à leur disposition de nouveaux objets "macro-phénomènes" pour enrichir leurs analyses exploratoires.

Dans une étape ultérieure, linguistes et informaticiens chercheront à étendre ce résultat à d'autres problématiques que la détection de séquences pour lever de nouveaux verrous scientifiques. Ces avancées pourront constituer les bases d'un projet

plus ambitieux comme la construction d'une ontologie dans le domaine de la gestion d'un conflit à l'oral.

Références

[1] Balthasar, L., Bert, M., (2005), *La plateforme « Corpus de langues parlées en interaction* (CLAPI) », *Lidil*, 31 .

[2] Dister, A., (2008), *L'autocorrection immédiate en français parlé : le cas des déterminants,* Université de Louvain – Centre de recherche VALIBEL, JADT 2008.

[3] Plantin, C. (2002): *Argumentation Studies and Discourse Analysis: The French Situation and Global perspectives Discourse studies 4,* 3.343-368.

[4] Malaisé, V., (2005), *Méthodologie linguistique et terminologique pour la structuration d'ontologies différentielles à partir de corpus textuels,* thèse de Doctorat de l'Université Paris 7 – Denis Diderot.

[5] Meurs, M-J., Duvert, F., Béchet, F., Lefèvre_renato, F., Demori, F., (2008), *Annotation en Frames Sémantiques du corpus de dialogue MEDIA* ,Université d'Avignon et des Pays de Vaucluse, TALN juin 2008.

[6] Eshkol, I., Tellier, I., Taalab,S., Billot,S., (2010), *Étiqueter un corpus oral par apprentissage automatique à l'aide de connaissanceslinguistiques,* JADT.

[7] Mokrane, A., Friburger, N., Antoine, J-Y., (2008), *Cascades de transducteurs pour le chunking de la parole conversationnelle : l'utilisation de la plateforme CasSys dans le projet EPAC,* Université François Rabelais Tours – LI, IUP Blois, France. TALN.

[8] Candéa, M. ,(2000), *Contribution à létude des pauses silencieuses et des phénomènes dits « dhésitation » en français oral spontané*. PhD thesis, Thèse de doctorat. Université Paris III.

[9] Blanc, O., Constant,M., Dister, M., Watrin, P., (2008), *Corpus oraux et chunking, (JEP)*.

[10] Yayoi, N-D., (2009), *Typologie des subordonnées et des connecteurs en vue de la détection automatique des propositions syntaxiques du français*, Description linguistique pour le traitement automatique du français, Cahiers du Cental.

[11] FORNEL, M., MARANDIN, J.-M., (1996). *L'analyse grammaticale des auto-réparations*. Le gré des langues 10, 8-68.

[12] Sauwens, C., (2000), *Etude de la dynamique d'autoassemblage chez plusieurs espèces de fourmis*. Thèse de doctorat, Université libre de Bruxelles.

[13] Hölldobler, B.,Wilson, EO., (1990). *The Ants*. Berlin: Springer Verlag.

[14] Blanche-Benveniste, C., Rouget Mireille Bilger, C., van den Eynde, K., *Le Français parlé,*(1990). Etudes grammaticales. CNRS Editions, Paris, Hershey, PA, USA.

[15] Valette,M, Garbar,N, (2004), *Caractérisation de texte à contenu idéologique: statistique textuelle ou extraction du syntagme,du projet pricipe.*

[16] Hearst, M., (1992), *Automatic acquisition of hyponyms from large text corpora*. In Proceedings of the 14th Conference on Computational Linguistics (COLING), pages 539–545.

[17] Paumier,s., (2004), *Unitex 1.2 Manuel d'utilisation*. Universite Marne-la-Vallèe, page54.

[18] Rizoiu, M-A., Velcin, J., Chauchat, J-H., (2010), *Regrouper les données textuelles et nommer les groupes à l'aide de classes recouvrantes*. EGC .

[19] Hajlaoui, K., (2008), *Dispositifs de recherche et de traitement de l'information en vue d'une aide à la constitution de réseaux d'entreprises*. Thèse de doctorat de l'Ecole Nationale Supérieure des Mines de Saint-Etienne,France.

[20] Blanc, O., Dister, A., (2004), *Automates lexicaux avec structure de traits*, RECITAL 2004.

[21] Laporte, E., Monceaux, A., (1997). *Grammatical disambiguation of French words using part of speech, inflectional features and lemma of words in the context*.

*[22] Juanals, B., Noyer, J.-M., (2007), D. H. Hymes, *vers une pragmatique et une anthropologie communicationnelle*. Laulan A.-M. et Perriault J. (dir.), Infocom : Réécrire la genèse. Revue Hermès CNRS, n° 47. Paris.

Summary

This paper presents a contribution which aims to automatically identify complex phenomena that make up an interaction like thanks, conflict, familiar jokes, etc. First, the phenomenon is studied, described and modelled on the basis of quantitative and qualitative analysis. Then, a set of tools for extracting knowledge will be designed to detect its instances automatically. The online database CLAPI (Corpus of Spoken Language in Interaction) that hosts recorded oral transcripts for different themes will be our corpus of data presented in XML format. We propose an approach for the identification of sequences and semantic annotation. This approach is based on methods of text mining and automatic processing of language. These results have been evaluated and validated by a linguistic expert in the field.

La terminologie des sciences humaines : le cas de la traductologie

Maha Kaddoura

Universite ParisIII - ESİT

Résumé. Notre travail entend révéler les particularités de la terminologie des sciences humaines en se basant sur un cas de figure : la traductologie. Il s'agit en l'occurrence de provoquer une sorte de dé-composition quasi organique de la « philosophie » de Wüster en vue d'une re- constitution ultérieure de ses composants fondamentaux au contact de la traductologie. C'est donc autour de ces composants décomposés : le concept, le terme et, notamment, l'objet, que s'articulera notre réflexion; cette tridimensionnalité de base sera scrutée à travers le prisme de la traductologie, une sorte de mise à l'épreuve, voire un affrontement entre présupposés classiques de la terminologie et nécessités de la traductologie. Cela permettra de mettre le doigt sur les divergences et les convergences, discerner les points communs et différents, dans le but ultime d'opérer une véritable translation comparative du concept, de l'objet et du terme vers un terrain quasi vierge pour la théorie terminologique : la traductologie comme discipline des sciences humaines.

1. Une rupture originelle

Depecker considère « la terminologie comme l'étude et le traitement des terme*s scientifiques ou techniques* »[1] (c'est nous qui soulignons). Or, si Depecker ne peut s'empêcher cette précision, ce n'est ni par hasard ni intentionnellement, mais en raison d'une volonté totalement inconsciente de la part des terminologues de rattacher leur discipline aux domaines techniques et scientifiques. Cette génétique scientifique a un tel caractère d'évidence en terminologie que Depecker le fait de manière tout à fait naturelle et, me semble-t-il, presqu'à son insu. Il est évident que « scientifique », entendu dans la signification qui motive la pratique terminologique, est rattachée à une action de la science visant l'objectivation de la Nature et non à toute action de pensée méthodique et critique. Par conséquent, il n'englobe pas ce que d'aucuns appellent les « sciences de l'humain » ou les « sciences humaines » mais se préoccupe uniquement des « sciences exactes ». En effet, personne ne peut nier que la terminologie a un passé fortement lié à la science et la technique, et ce malgré toutes les formules de distanciation théorique – et purement rhétorique – que

1 DEPECKER, Loïc, « La terminologie est-elle une science ? », in *La terminologie, discipline scientifique*, actes du colloque du 17 octobre 2003, organisé à l'ENS de la rue d'Ulm, Ed. Société française de terminologie, p.11.

les terminologues emploient dans leurs écrits. Maria-Teresa Cabre le dit d'ailleurs clairement quand elle cite l'évolution de la science et de la technique comme un des facteurs qui ont contribué a la création et l'évolution de cette discipline[2]. Certains terminologues affichent néanmoins la volonté de libérer la terminologie de ces limites réductrices en essayant de lui donner une portée beaucoup plus générale. Pour clarifier notre propos, nous avons jugé pertinent d'étudier la définition que Guy Rondeau donne aux termes scientifique et technique: « Il faut entendre « scientifique » au sens le plus large incluant les sciences pures et appliquées, les sciences connues sous le nom d'*exactes, les scie*nces humaines, etc. ; de même, « technique » recouvre les technologies, les techniques, les activités spécialisées (métiers, professions, occupations, etc.) de travail ou de loisirs. »[3] (c'est nous qui soulignons) Cependant, cet auteur demeure tout au long de son ouvrage incapable de présenter ne serait-ce qu'un seul cas issu des « sciences humaines » et les rares exemples qu'il donne pour étayer ses thèses sont empruntés à des domaines tels la médecine, l'industrie de l'imprimerie, l'informatique, la chimie, la technique (fonctionnement de machine) ou concernent des objets concrets du monde (ex : coffee mill/moulin à café à coteaux).

Ce schéma conflictuel qui oppose une volonté de généralisation à une application limitée met en jeu une théorie surdimensionnée par rapport à sa praxis restreinte. Ceci s'applique également à l'emploi de l'adjectif « technique ». A la base difficile à cerner, cet adjectif est surtout utilisé en terminologie dans deux significations subtilement évoquées par le dictionnaire général, le Petit Robert : « qui concerne les applications de la connaissance théorique, dans le domaine de la production et de l'économie » et « qui concerne les objets, les mécanismes nécessaires à une action ». Or certains terminologues se refusent à limiter ainsi la signification de « technique » et tentent, à l'instar de Robert Dubuc, d'élargir sa sphère sémantique en affirmant que « le mot technique recouvre la quasi-totalité de l'agir humain, à la condition qu'il ait fait l'objet d'une certaine codification. Il englobe donc les arts, les sciences, les métiers et les diverses branches de l'industrie ainsi que certaines activités de loisir comme le sport. »[4] A première vue, cette définition confère à ce terme un véritable caractère de généralité (« quasi-totalité de l'agir humain ») et pourrait paraître contredire nos déductions précédentes, mais il suffit de regarder plus attentivement pour remarquer que, sous une apparence d'élargissement sémantique,

---

2 CABRE, Maria Teresa, *La terminolo*gie, théorie, méthode et applications, Traduit par Monique CORMIER et John HUMBLEY, Ed. Les Presses de l'Université d'Ottawa, Ottawa, p.40.

3 RONDEAU, Guy, Introduction à la terminologie, Ed. Gaëtan Morin, 2ème édition, Québec, 1981/1984, p.13

4 DUBUC, Robert, *Manuel pratique de terminologie*, Ed. Linguatec, 4ème édition, Québec, 1930/2002, p.6.

cette définition n'est en réalité qu'une tentative artificielle, parce que purement théorique, d'élargir le domaine d'application de l'adjectif « technique ». En effet, usant de termes vagues et ambigus tels « arts », « sciences », « agir humain » et « codification », sans prendre le temps de les expliquer, ni dans le cadre de sa définition ni ultérieurement dans son ouvrage, Dubuc alourdit sa définition par une sorte d'accumulation de généralisations conférant un semblant de globalité à cet adjectif pour tenter d'échapper à la restriction sémantique qu'il laisse supposer. Dubuc d'ailleurs se trahit lui- même quelques pages plus tard dans un surprenant glissement sémantique quand il dit : « Ainsi la qualité d'extrême résistance à l'usure sera désignée différemment selon *qu'il s'agira d'un discours publi*citaire ou technique, ou encore si elle est attribuée à un produit de consommation ou à un bien d'équipement. »[5] (c'est nous qui soulignons) « Technique » est vraisemblablement utilisé ici dans une des deux significations que donne le Petit Robert et prouve bien que la définition de Dubuc est coupée de la réalité de la terminologie telle qu'elle se pratique. Et c'est justement cette rupture entre ce genre de formules définitoires, si saisissantes soient-elles, posées par les auteurs au début de leur ouvrage, qui veulent coûte que coûte donner un semblant d'universalité à leur discipline, et l'utilisation ultérieure, souvent contradictoire, qu'ils font de «scientifique» ou «technique», qui laisse déjà entrevoir ce rapport extrêmement pesant des sciences et techniques sur la terminologie. C'est dans la difficulté d'appliquer la théorie à la pratique que se manifeste la difficulté de la terminologie à se libérer de son passé. C'est cette théorie constamment mise en doute par la pratique qui accrédite davantage la pratique et affaiblit la théorie.

Or, le schéma classique du triangle sémiotique (signe-concept-objet) ne laisse en rien supposer un quelconque lien entre la théorie de la terminologie et l'esprit scientifique ; s'il semble avoir une visée plutôt générale, c'est en se plongeant davantage dans l'analyse de chaque pôle du triangle sémiotique, dans les relations qu'ils entretiennent entre eux et dans la portée pratique de ce schéma qu'apparaît toute cette dimension cachée. La vocation scientiste de cette discipline n'est pas directe, elle n'est pas revendiquée clairement par Wüster, elle transparaît dans la manière de concevoir le concept, le signe et l'objet et surtout dans la manière de concevoir les relations qui lient ces trois pôles. Tout d'abord, en terminologie, l'objet semble posé dans le monde, facilement livré à la recherche qui est là pour l'étudier

dans sa concrétude immédiate. Autrement dit, l'objet du triangle terminologique est appréhendé, semble-t-il, comme un référent spatio-temporel, u*n e tre* matériel, comme une unité donnée dans un monde à référence objective, un objectum. Or cette vision empirico- matérialiste de la réalité des objets est en relation étroite avec les réalités physiques et biologiques qui étudient les objets concrets du monde. La dimension concrète indéniable de l'objet lui permet d'asseoir toute la stabilité

5 DUBUC, Robert, Ibidem, p.13.

ontologique de cette discipline. On ne peut rien comprendre à la précision et la concision que revendique la terminologie si l'on ne prend pas en compte cette présence de l'objet dans un monde sensible qui nous est directement accessible. Par ailleurs, la solidité ontologique de la terminologie ne peut que se refléter sur le concept ; elle lui offre une stabilité. Car le concept permet d'appréhender les caractéristiques objectives qui font qu'une chose est ce qu'elle est. Assis sur une base ontologique inébranlable, le concept se voit attribuer des qualités d'immuabilité et de stabilité absolues. Or, une telle stabilité conceptuelle ne peut être envisagée sans l'idée d'une certaine pureté transcendantale où le concept est en quelque sorte rationalisé au plus haut point, détaché de tout ancrage sociologique. C'est dans cette conceptualisation épurée que l'on retrouve

pleinement l'idée de l'universalisation des concepts : du moment que les concepts sont élevés au stade d'idéalité conceptuelle pure, ils deviennent potentiellement universels, voire universalisables. L'idée même de catégorisation et de classement du savoir provient d'une tradition naturaliste et rentre bien dans la lignée de la scientificité. Et le signe ? En terminologie, comme nous le savons, le concept est caractérisé par un ensemble d'attributs immuables qui permettent de le reconnaître dans tout corpus textuel. D'un autre côté, le terme (ou signe) est caractérisé lui par son association à un ensemble d'attributs qui, eux, permettent de définir le concept que le terme désigne à l'exclusion de tout autre concept. Cette relation d'équilibre réciproque instaure une certaine stabilité entre le concept et le terme. C'est particulièrement grâce au principe d'uninotionnalité que cette stabilité est préservée en terminologie car son absence cause le "flou cognitif " de la synonymie, de la polysémie, de l'homonymie et de tout autre phénomène d'ambiguïté conceptuelle.

Mais de cette relation très forte entre concept et terme, se dégage paradoxalement une rupture fondamentale entre le conceptuel et le linguistique où le conceptuel prime sur le linguistique. Le signe apparaît pour ainsi dire comme une étiquette posé sur un concept pur et transcendantal. Le scientisme se dote d'une dimension idéaliste qui n'est pas sans rappeler les présupposés platoniciens. Dans cette optique, il existe un monde des idées (les concepts) qui correspond point par point à un monde sensible (les objets) et qui sont dotés de noms (les termes) afin de les reconnaître. A la vision scientifique de la terminologie, s'ajoutent des résonances platoniciennes : ce n'est donc pas n'importe quel scientisme qui sous-tend la terminologie, c'est un idéalisme scientifique. Or peut-on transposer une telle vision du monde, génétiquement scientifique et théoriquement idéaliste, à une discipline de nature profondément instable, en l'occurrence la traductologie ?

Dans sa quête de stabilité ontologique et de précision conceptuelle, la terminologie fait figure de discipline on ne peut plus idéale pour donner aux domaines spécialisés qui y ont recours la possibilité d'organiser leurs termes et de fixer leurs concepts. Véritable pourvoyeuse de solutions efficaces à des problèmes

précis, la terminologie pourrait paraître comme «la» réponse aux dérangements terminologiques qui perturbent les disciplines spécialisées. Pourtant, en regard de cette prétention à la stabilité, l'acte fondateur de la terminologie, à savoir le triangle sémiotique, est profondément mis à mal tant par la traductologie que par la majeure partie des sciences humaines. Serait-ce son extraordinaire puissance qui s'assimile tout d'un coup à son extrême faiblesse ? Déstabilisée dans son propre pouvoir de stabilisation, la terminologie semble avoir un édifice théorique qui ne résiste pas à la traductologie. Car traductologie et science exacte sont des vertus d'ordre différent et en négligeant cette importante différence, on ouvre évidemment la porte aux plus funestes aberrations. Toute la difficulté réside dans l'application de la terminologie à un domaine qui n'a pas forcément les mêmes tenants et aboutissants de ses domaines de prédilection. Il convient de s'avancer très prudemment sur ce nouveau terrain glissant où les limites sont loin d'être étanches et où toute tentative de fixation n'est qu'un essoufflement semble-t-il inutile. La question qui se pose est évidemment pourquoi ? Pourquoi la terminologie ne peut s'appliquer à la traductologie et, presque tautologiquement, pourquoi la traductologie se dérobe-t-elle à toute application terminologique ? Nous nous garderons bien de faire un exposé sans fin sur les divergences entre les sciences humaines et les sciences exactes qui risquerait d'acheminer le sujet vers un dédale de théories philosophiques. Nous avons plutôt décidé, tout en évitant une caricature trop réductrice, de faire pivoter le sujet autour du triangle sémiotique de la terminologie – et d'entamer sa dé-composition. Commençons tout d'abord par effectuer une sorte de comparaison, qui ne peut qu'engendrer une confrontation, entre l'objet du triangle sémiotique tel que le perçoit la terminologie, et l'objet d'étude de la traductologie. Nous glisserons par la suite dans cette comparaison les éléments pertinents relatifs aux concepts et aux termes. Comme nous l'avons mentionné, l'objet de la terminologie est une unité spatio-temporelle matérielle. Si la préoccupation de la terminologie de bien cerner les concepts apparaît possible, c'est en premier lieu grâce à la concrétude des objets sur lesquelles elle se penche, et c'est parce que les sciences et les techniques e ***voluent dans*** un univers ***d'objets maj***oritairement ***concrets, ou quantifiab***les, qu'elles se prêtent facilement à une étude terminologique. Or peut-on penser tran***scodage, ambi***guïté, vouloir- dire, fidélité, traductologie inductive en termes *spatio*-temporels matériels ? Evidemment non. Car comment relier par exemple le vouloir-dire à un objet concret s'il est par définition

OBJOBJ

un "état de conscience pré-verbal qui entraîne de façon nécessaire l'émission de paroles. [...] »[6]. (c'est nous qui soulignons) Cet état de pré-verbalisme enlève d'emblée toute concrétude potentielle. Il n'est certainement rien de « matériel » dans

6 Glossaire in LEDERER, Marianne, La traduction aujourd'hui: le modèle interprétatif, Ed. Hachette, Paris, 1994, p.218.

cette unité-là, mais aussi rien de circonscrit, d'enclavé dans une sorte d'enceinte spatio-temporelle rigide. En traductologie, ce que l'on cherche à accorder ce sont des rapports de langue, de pensée, de culture, qui n'ont pas de volume saisissable, mais qui se dévoilent dans un univers qui, s'il s'inscrit dans le temps et l'espace, ne suit pas pour autant leur cours et contours normatifs. Il n'est pas d'univers matériel qui fonde les références traductologiques. Et la réalité traductologique n'est nullement rattachée à une théorie de la référence matérielle qui relie les termes à des objets concrets du monde extra-linguistique, théorie incapable de relever ce qu'il peut y avoir de plus insaisissable dans l'expérience traductologique. Les Vérités traductologiques n'apparaissent donc pas dans la concrétude de la réalité mais dans une abstraction de l'expérience linguistique, culturelle et humaine. Dans son mouvement pensant, son avancement spéculatif, la traductologie s'édifie sur des allers retours permanents entre ce qu'il y a d'intangible dans l'expérience de la langue et celle de la culture, agençant un orchestre d'éléments multi-disciplinaires dans une sorte d'abstraction de l'expérience humaine par excellence. Ce n'est pas une abstraction axiomatique, objectivement mesurable, telle que l'abstraction mathématique, mais une abstraction inquantifiable. C'est ainsi que la traductologie a certainement un objet de référence mais son référent n'est pas celui étudié en terminologie. Il est abstrait. Et cette ontologie abstraite de la traductologie s'accomode très mal de l'ontologie concrète de la terminologie. La terminologie se heurte ainsi à une discipline où la spéculation fait figure de pensée ambiguë, flottante. On marque ici une nette différenciation entre la traductologie et la terminologie, comme si le terme « abstraction » reprenait son sens étymologique et devenait par essence même séparation, isolement. Au-delà d'une différence purement contingente, l'objet de la traductologie apparaît pour ainsi dire en contradiction fondamentale, même substantielle, avec l'objet de la terminologie. C'est ainsi qu'au contact de la traductologie, l'« objet » du triangle sémiotique est symboliquement supprimé. Privé de sa source de stabilité, le triangle sémiotique est amputé, laissant pour ainsi dire la traductologie dans un état de flottement ontologique. Sans point d'appui concret ou quantifiable, sans réalité « sensiblement » dressée, la traductologie se déploie dans un abstrait toujours recherché mais jamais donné. Le sens ne nous a toujours pas livré tous ses secrets et ne le fera probablement jamais, le thème de l'altérité est loin d'avoir épuisé toutes ses ressources. Contrairement aux sciences exactes, qui se livrent dans une concrétude directe, dans la brutalité de la chose livrée, la traductologie ne se dévoile que lentement, indirectement et progressivement au sujet connaissant. Les sciences exactes ont une relation d'être ou de néant, tandis que la traductologie est souvent dans l'étant. C'est la raison pour laquelle les concepts des sciences exactes sont tels quels, exacts ou inexacts. Ils demeurent ou s'évanouissent. Le principe en traductologie est différent, les concepts sont loin d'être dans une dimension manichéenne de vie ou de mort, ils sont dans une logique d'évolution permanente, de reprises, de rectifications. Il n'y a pas la brutalité des sciences exactes parce qu'il n'y a pas la brutalité de leurs inventions et innovations. En traductologie, le désir de

s'approprier l'objet est éternellement inassouvi si bien que l'analyse est condamnée à varier sans relâche les angles de vue pour cerner un objet toujours incernable. Celle-ci pourrait donc être indéfiniment reprise. Quand l'objet d'étude investit la langue, voire la parole, avec tout ce que cela implique aux niveaux culturel, cognitif et communicationnel, on voit bien que la tâche n'est pas facile et que la Vérité des représentations qui sont en jeu en traduction est très difficile à épuiser. Et c'est probablement dans cet univers abstrait – qui ne se donne à aucune vérification, dans ce champ de vérité insaisissable, partielle, que l'on peut voir toute la dimension multiple et subjective de la discipline. L'on remarque ici, au passage, une dimension en contradiction flagrante avec la terminologie. Plus précisément, l'objectivité que prône la terminologie s'oppose à la multiplicité subjectivante de la traductologie. En effet, chaque penseur, chaque chercheur, tente à sa manière, de par sa propre expérience, en fonction de son point de vue, de tracer un chemin qui mène à une part de vérité traductologique. Il suf***fit de vo***ir toutes les significations attribuées à contexte pour mesurer la diversité des angles de recherc*hes possibles. Nous n'en citerons* que deux : celle de La Terminologie de la Traduction qui pose la définition suivante: "Environnement linguistique d'une unité lexicale qui contribue à préciser la ***significa***tion pertinente en vue de dégager le sens." [7] Le contexte y est donc limité à la seule variante linguistique du texte. Marianne Lederer, par contre, donne une définition tout à fait contraire en disant "[...] contexte désigne les circonstances qui entourent un texte, de près (circonstances d'émission d'un discours) ou de loin (ensemble de la situation historique, sociale, économique mais aussi personnelle dans laquelle ce texte a vu le jour). Il s'agit de l'entourage non linguistique dans lequel un énoncé est produit ou reçu."[8] C'est incontestablement une multiplicité de virtualités qui s'entremêlent, de possibles sans jamais être. Il y a une participation active du sujet à l'objet. Les sciences exactes ont pour objet des faits qui se présentent à la conscience comme des phénomènes donnés isolément de l'extérieur, tandis que les références de la traductologie se présentent au sujet de l'intérieur comme un ensemble de virtualités. Il s'agit en l'occurrence de mettre en lumière le côté subjectif de la pensée traductologique. En effet, en traductologie, le sujet ne peut se détacher de l'objet, expurgeant la connaissance de tout ce qui, en elle, est tributaire de l'existence humaine. La traductologie, comme beaucoup de sciences humaines, prend donc une allure subjectivante qui la rend par définition même instable, mouvante, non quantifiable, et par-dessus tout livrée aux différents angles d'étude, parfois complémentaires et souvent contradictoires, des traductologues. On voit bien la dimension flottante immanente à la traductologie et par conséquent incompatible avec la terminologie. Entre l'ambivalence abstraite, le doute

---

7 DELISLE, Jean et al, *Terminologie de la traduction*, Coll. FIT, Ed. John Benjamin, Amsterdam/Philadelphia, 1999, p.22.

8 Glossaire in LEDERER, Marianne, La traduction aujourd'hui: le modèle interprétatif, Ed. Hachette, Paris, 1994, p.212.

ontologique de la première et la concrétude cartésienne, la certitude positiviste de la seconde, il y a un véritable fossé épistémologique et la terminologie a donc pendant longtemps rechigné à s'attaquer aux sciences humaines. En effet, prônant la stabilité, la précision et l'objectivité, la terminologie est complètement désemparée face à une discipline de nature mouvante qui est, par essence, instabilité, ambiguïté et subjectivité. Privée d'un ancrage profond dans la réalité matérielle et d'une logique classificatrice rigoureuse, la traductologie n'a pas les moyens d'adopter une vision du concept pur et clair et de se doter d'un lien de stabilité et d'univocité absolues entre le concept et son terme. C'est une des raisons pour lesquelles on ne peut appliquer à la traductologie une étude logique rigoureuse de ses termes et concepts. Même d'un point de vue purement formel, on trouve rarement des unités terminologiques qui peuvent être décelées grâce à leurs *composan*ts formels (souvent greco-latin).

Evoluant souvent dans l'indicible, les concepts de la traductologie deviennent, par définition, ce « qu'on ne peut caractériser par le langage ». Comment alors réaliser dans l'ineffable le but ultime de la terminologie qui est d'attribuer une définition à un terme ou d'assigner un terme précis à un concept ? Comment concilier l'inconciliable ? Et c'est là que se déploie une grande part de la relation tragique qui noue – ou dénoue – la terminologie à la traductologie. En effet, il découle de notre analyse que la difficulté de la traductologie à se soumettre à une étude terminologique rigoureuse relève en partie de l'absence d'un concept qui se livrerait entièrement dans une mise en forme linguistique. Il reste toujours une part de non-dit, d'indéfinissable. Les réalités traductologiques ne pouvant jamais être vérifiées de manière incontestable laissent toujours le terminologue ou le traducteur dans un état de frustration où la signification des concepts lui échappe perpétuellement. Qu'est-ce que l'éclairage ? Comment ***mesur***er la perte dans une traduction ? Commen***t e valuer***l'intention d'un auteur ? Les concepts sont suffisamment abstraits et ont assez d'implications pour générer des conflits parmi les spécialistes à propos de leur contenu, et même le choix du terme pose souvent des difficultés. Par ailleurs, souvent la définition d'un terme en traduction s'appuie sur des présupposés conceptuels qui sont souvent eux-mêmes loin d'être clairement identifiés et cernés. Pour prendre un exemple anglais, citons la définition que Peter Newmark donne de la communicative translation. C'est une traduction où "the translator attempts to produce the same effect on the TL readers as was produced by the original on the SL readers."[9] (c'est nous qui soulignons) Mais qu'est ce que l'effet? Comment se mesure t-il ? Peut-il être mesuré ? Il y a d'emblée une foule de questions qui assaillent le lecteur et rendent la compréhension de ce terme inéluctablement liée à la signification d'un autre terme en soi difficilement

9 NEWMARK, Peter, *A Textbook of Translation*, Ed. Prentice Hall, Hemel Hempstead, 1981/1998, p.22.

compréhensible. Face à la précision des termes dans les domaines scientifiques et techniques, nous retrouvons cette impression de flottement conceptuel, comme une pensée en attente perpétuelle d'accomplissement, de complétude. Les conséquences se voient directement: polysémies, synonymies etc., la traductologie foisonne d'irrégularités conceptuelles au point où elle échapperait à toute tentative de fixation. Car toute définition posée qui fixe les limites sémantiques d'un terme est vouée à un statut de vérité provisoire, personnelle, partiale. La mise en forme linguistique pourrait même être perçue comme une tentative de stabiliser des concepts qui sont par nature évanescents, de donner une seule facette à un objet qui pourrait par essence en avoir plusieurs ; c'est une tentative de maîtriser linguistiquement un concept qui se refusera souvent, même dans son moule linguistique, à livrer toute sa dimension. Et c'est dans ce mouvement perpétuel, cette tentative d'évasion du concept de la forme dont il est captif que la traductologie rentre en collision avec la terminologie qui est essentiellement prescriptive et normative. En fait, la terminologie qui a pour fondement la normativité et pour but suprême la normalisation se voit dans ce cas-là devant une impossibilité à fixer des normes. Josette Rey-Debove a d'ailleurs raison lorsqu'elle dit que les sciences humaines sont dans un état d'impunité par rapport aux autres sciences. "On ajoutera, et ceci n'est pas sans rapport, que les chercheurs en sciences humaines

sont mis dans une situation d'irresponsabilités par l'absence de sanction [...]."[10] En effet, en revenant à l'exemple de la définition ***de conte***xte, qui peut juger si c'est *le Dictionnaire de la Terminolog*ie de Jean Delisle qui a raison ou le glossaire de Marianne Lederer? Personne. Parce que personne n'a « ontologiquement » raison. Qui pourrait dire que tel terme est juste ou tel autre ne l'est pas ? Personne non plus. Comment alors parvenir à stabiliser des concepts qui par défaut échappent à toute systématisation si ce n'est probablement au prix de quelque déformation, concession ou même renoncement conceptuel ? L'application que nous avons commencé à développer est certes, modeste, mais ce premier travail a permis d'affiner les approches et d'entrevoir les difficultés dont on perçoit mieux la nature et l'ampleur. Les questions qui se posent en conclusion ouvrent le débat sur une éventuelle solution à cette antinomie originelle entre terminologie et traductologie. Doit-on trouver une autre discipline plus apte à dialoguer avec la traductologie, ou plutôt essayer d'appliquer la terminologie quitte à échouer ou à la voir complètement dénaturée ? Car, utiliser la terminologie dans un domaine des sciences humaines, c'est en quelque sorte l'arracher à ses origines, lui attribuer une part d'humain qu'elle n'avait pas au début. La tentation n'est-elle pas de concéder aux sciences exac*tes l'autorit*é en matière de factum rationis et de reléguer les sciences humaines

---

10 REY-DEBOYE, Josette, "Spécificité de la terminologie linguistique", in Métalangage et terminologie linguistique, Tome I, Actes du colloque international de Grenoble – Université de Stendhal – Grenoble III, 14-16 mai 1998, Ed. Peeters, Coll. Orbis Suppementa, tome 17, Leuven/Paris/Sterling/Virginia, 2001, p.5.

au rang de « sciences secondes », consommant ainsi la rupture originelle entre ces deux disciplines ?

2. Rupture finale

Terminologie et traductologie sont donc incompatibles. C'est la rupture originelle, un constat d'échec mutuel. Car si la traductologie apparaît effectivement instable et floue, la terminologie n'en demeure pas moins partielle. Cette lecture presque bijective qui met ces deux disciplines en confrontation soulève deux questions : La traductologie est-elle vouée à une sorte de verbalisme chaotique, incapable de pensée exigeante? Tandis que la terminologie reste, elle, de son côté, l'apanage exclusif des domaines technoscientifiques ? Ce n'est pas ce que donne à penser l'histoire qui montre sans équivoque que la langue dite « savante » – qui véhicule les savoirs – formée en Europe entre le Ve et le XIXe siècles, ne s'est pas penchée exclusivement sur des domaines *scientifiques* stricto sensu ; la philosophie et la sociologie y avaient déjà une place prépondérante. Historiquement, la terminologie se trouve donc liée à l'éclosion d'un certain nombre de sciences humaines, surtout au moment où les langues vernaculaires envisageaient de se doter d'un arsenal terminologique leur permettant de dépasser la suprématie de la langue latine. Par ailleurs, et en se plaçant dans un contexte bien plus récent, il ressort de l'observation terminologique des sciences humaines une conclusion pour le moins qu'on puisse dire intéressante : beaucoup de spécialistes essaient à leur façon d'étudier la terminologie de leur discipline. Ce n'est pas tant l'application de présupposés théoriques rigoureux qui les motive que l'organisation des termes, la clarification des concepts de leur domaine, etc. En traductologie par exemple, et nous en avons déjà parlé, le *dictionnaire de la Terminologi*e de la *traduction* [11]*, le Dictionary of* Translati*on Studies*[12] *et le Routledge Encyclopedia of Tra*nslation Studies[13] ont été produits. Donc malgré le renoncement de la terminologie aux sciences humaines, la traductologie entre autres a plus ou moins pris conscience, par intuition ou par besoin, de la nécessité d'étudier ce qu'elle considère comme ses termes. Sur quoi les uns diront que c'est une tentative vaine, les autres que c'est indispensable. Le fait reste que, dépassant le pur soubassement théorique de la terminologie, les traductologues ont conduit tant bien que mal, une entreprise visant

---

11 DELISLE, Jean et Hannelore LEE-JAHNKE, *Terminologie de la traduction*, Coll. FIT, Ed. John Benjamin, Amsterdam/Philadelphia, 1999, 433p.

12 SHUTTLEWORTH, Mark et Moira COWIE, *Dictionary of Translation Studies*, Ed. St Jérome Publishing, Manchester, 1997.

13 BAKER, Mona, *Routledge Encyclopedia of Translation Studies*, Ed. Routledge, Taylor and Francis Group, New York/London, 1998 (1ère éd.), 2001, 654p.

à organiser leurs termes. Pourrait-on en conclure que la terminologie renferme en elle des possibilités d'application ignorées, un potentiel inconscient qu'il faudrait interroger davantage, à la manière d'une maïeutique socratique, pour le pousser à se dévoiler ?

Si tranchée que paraît à première vue la distinction entre terminologie et traductologie, une étude plus attentive ne tarde pas à l'atténuer considérablement. En effet, comme nous l'avons montré, l'intuition du spécialiste vient tempérer le dogmatisme du terminologue qui, en retour, ne tarde pas à modifier le regard qu'il porte sur sa discipline. Alain Rey, par exemple – et l'on ne peut qu'approuver sa démarche – s'engage dans une critique tout à fait pertinente s'agissant de la dérive scientiste qui a radicalisé la terminologie. Selon lui, il est difficile de transposer des méthodes de travail relatives aux « sciences plus formalisées » sur ce qu'il nomme « des savoirs plus incertains » ; et de rajouter que « le positivisme rigide qui gouverne souvent la théorie terminologique est probablement une étape nécessaire, au moment de sa constitution. Mais le décalage entre la volonté de description et la méconnaissance des objets à décrire condamnerait cette activité à l'impuissance (ou au statut modeste d'aide à la traduction, à la documentation), alors qu'il s'agit véritablement d'un corps de connaissances indispensables pour satisfaire à un besoin fondamental préalable à tout réglage social. » Attestant du rôle du positivisme dans l'avènement de la théorie terminologique, Rey en relève pourtant les limites : la méconnaissance de ce fameux « objet » à décrire. Méconnaissance qui, sans s'y fondre complètement, évoque l'objet des sciences humaines « méconnu » - qui n'est ni connu ni apprécié à sa juste valeur - de la terminologie. En fait, Rey sent bien le réductionnisme de la vision terminologique classique et laisse entendre la nécessité d'aller au-delà du « positivisme rigide » tel qu'il le nomme. Or, « méconnaître » une chose, c'est aussi ne pas la connaître pour ce qu'elle est ; pris dans cette signification, ce verbe n'entend pas nier en bloc la terminologie dans son approche positiviste classique, mais plutôt combler la vision en partie ignorante de cette approche. Il semble que Rey veuille que cette discipline comprenne son objet d'étude dans toute sa plénitude, probablement afin de rétablir la cohésion d'une discipline dont la prétention généralisatrice se voit souvent trahie par la pratique. Ces remarques font surgir et permettent de pointer à juste titre le problème de la radicalité de la terminologie classique : un fondamentalisme absolu qui ne donne qu'une lecture radicale du réel. Et c'est cette radicalité qui pousse maints linguistes et terminologues à se rallier au point de vue d'Alain Rey et à prendre conscience de la nécessité d'une pondération de la terminologie. Le positivisme en tant que doctrine, et dans son rapport à la terminologie, est largement remis en cause. Et l'on voit apparaître, après les dérives d'une scientifisation positiviste incontrôlée, un courant qui incarne de par son nom même l'*infil*tration du social en milieu terminologique : la socioterminologie.

La socioterminologie part donc du principe qu'il n'y a aucune évidence ontologique et conceptuelle dans la formation et l'utilisation des termes, le social ne

pouvant être négligé comme facteur d'influence primordial. Et ce qui est extérieur à la terminologie, à savoir les conditions de production et de diffusion, résonne forcément à l'intérieur de cette discipline, au sein du triangle sémiotique. Donc pas de cloison. On assiste à l'« intrusion » d'éléments sociaux qui réagencent les relations objet-signe-concept. Réagencement neutre ? Evidemment non. Car il est provoqué par un élément diversiforme, la société. Fortement compromis par cette intrusion, le triangle ne reste plus fidèle à ses postulats de départ et voit la stabilité absolue dont il faisait preuve très perturbée. De fait, l'univocité conceptuelle cesse d'être un gage de stabilité pour se transformer en une « obsession fétichiste »[14] qui fige des rapports par essence dynamiques. Reléguée au rang de « rêve » ou d' « idéal », l'univocité conceptuelle apparaît comme porteuse de frustration car vouée à ne jamais voir le jour ; la rigueur des définitions, le consensus absolu sur les significations, l'utilisation univoque des termes... sont perpétuellement contredits par la pratique. Selon Yves Gambier, « le consensus ne peut être que partiel (il ne couvre jamais tous les locuteurs ni toutes les notions du domaine) »[15]. Aussi le regard porté sur ce qu'on appelle « exceptions » ou «variantes» change-t-il considérablement. Les faits de synonymie, polysémie, ambiguïté conceptuelle... apparaissent inévitables et ne sont plus à bannir à tout prix. Dans cette perspective, la conceptualisation est concessions et consensus, la terminologie négociée devient inévitable. La complexité des usages mis en jeu exige de décloisonner la vision trop classificatrice des vocabulaires spécialisés : la catégorisation trop rigide des concepts des domaines est à remettre en cause si la théorie terminologique veut correspondre au dynamisme de la réalité terminologique, porteuse de multiplicité, d'interactions, d'apories, d'entente partielle et de fixations provisoires. De cette remise en cause, l'on tire la conclusion suivante : ce n'est visiblement pas en sclérosant la vision de l'objet, du concept et du terme que la terminologie peut réellement s'enrichir, mais en reprenant sa légitimité a*u* sein de la langue naturelle et en *se r*estitu*ant sa par*t de social. L'esse et le conceptum de la terminologie classique sont mis à rude épreuve.

Cependant, il ne s'agit nullement d'une métamorphose de la terminologie, mais d'une sorte de remodelage du triangle sémiotique. En effet, la terminologie continue à se baser sur ce triangle tripolaire (concept-signe-objet). Son existence n'est en aucun cas remise en cause mais les relations entre les trois pôles ne sont plus issues d'un scientisme idéaliste. Toutefois, parce que ce remodelage socioterminologique concerne uniquement les termes technoscientifiques, il n'a pas de vocation globale. Dans son célèbre ouvrage fondateur[16], François Gaudin, développe des idées qui s'attachent aux sphères d'activité proches des sciences. Ses successeurs ne dérogent

14 GAMBIER, Yves, "Présupposés de la terminologie: vers une remise en cause", i*n Terminologie et sociolinguistiq*ue, édité par François GAUDIN et Allal ASSAL, Coll. Cahiers de linguistique sociale, N°18, 1991, p.42.

15 GAMBIER, Yves, Ibidem, p.45.

pas à la règle. Et l'on voit ici en quoi ce changement n'a rien d'une métamorphose. Cette constatation est déjà bien formulée pour nous inquiéter sur l'aptitude de la terminologie à vraiment s'appliquer à la traductologie. C'est cependant s'arrêter à une constatation de facilité intellectuelle dans laquelle s'est justement complu la terminologie. En fait, même si les sciences humaines sont exclues de la réflexion socioterminologique proprement dite, il n'en reste pas moins que cette vision d'une terminologie plus nuancée est bien plus apte à accueillir la composante humaine de la traductologie. Comme l'on y accorde plus d'importance à la dimension socio-discursive des termes, on accepte d'emblée qu'il peut y avoir des limites non étanches et des réalités non univoques. C'est déjà une ouverture inédite à la nature profondément mouvante et multiple de la traductologie. Il y a là un*e esquisse d'*hominisation, pourrait-on dire, permettant à la terminologie d'opter pour des choix théoriques moins tranchés et d'intégrer progressivement cette part de subjectivité qui caractérise la traductologie. Où est donc le problème ? Ce remodelage part toujours du modèle de Wüster. On reste dans l'esprit de son schéma auquel on rajoute des nuances humaines et sociales. Il est vrai que l'on marque un certain écart par rapport à la théorie classique, mais l'on se glisse toujours dans ses cadres déjà constitués, ses circonscriptions déjà définies. C'est-à-dire que l'on critique effectivement, comme nous en avons parlé, l'idéalisation, l'univocité conceptuelle, la vision insulaire et statique du terme... mais l'on reste très souvent dans la « valorisation » de ces préceptes. Guespin dit : « La tendance à l'univocité reste la règle même si elle ne réalise jamais son idéal [...] »[17] Autrement dit, la socioterminologie ne renie *p*as la conception classique en tant qu'idéal, mais juste en tant que réalité. Ce qu'elle veut, c'est pousser la terminologie à se rende compte que ce qu'elle pense être une réalité accessible, est en fait un idéal que nul terme ne peut atteindre mais auquel il peut certainement continuer d'aspirer. Le « modèle » wüstérien est donc à prendre au pied de la lettre. Mais est-ce vraiment un idéal à atteindre – si idéal il y a - ou est- ce tout simplement une illusion en rupture totale avec la réalité ? Pour l'instant, l'on notera que le social prend une place dans l'étude des termes mais est toujours considéré comme un facteur d'instabilité face à une volonté terminologique stabilisatrice. Cette idée transparaît dans les articles de traducteurs – car c'est surtout eux qui le font – portant sur la traduction spécialisée en sciences humaines. Ils énumèrent les problèmes spécifiques à la traduction des termes de ces domaines mais ne peuvent s'empêcher de considérer la terminologie presque comme un remède. Arilla Escartin, une des rares traductrices qui s'est penchée sur la traduction des termes de l'Histoire dit : « Dans la décision de la [un exemple d'unité

16 GAUDIN, François, Socioterminologie, une approche sociolinguistique de la terminologie, Ed. Duculot De Boeck, Bruxelles, 2003, 286p.

17 GUESPIN, Louis, « La circulation terminologique et les rapports entre science, technique et production, in *Terminologie et sociolinguistique*, édité par François GAUDIN et Allal ASSAL, Coll. Cahiers de linguistique sociale, N°18, 1991, p.68.

terminologique] faire entrer dans notre éventuel fichier terminologique, le facteur subjectif joue un rôle indéniable, mais il est quand même, précisément pour cette raison, indispensable d'établir des critères préalables qui dirigent la sélection des termes significatifs. »[18] (c'est nous qui soulignons) Arrêtons-nous tout d'abord à l'expression « précisément pour cette raison ». L'auteur ne nie pas le rôle de la subjectivité dans la formation des termes de l'Histoire, mais d'un autre côté c'est cette subjectivité – à entendre cette instabilité – qui interpelle davantage la nécessité d'un travail terminologique organisateur et normalisateur («critères préalables »). Persiste donc cette illusion normative d'une terminologie en mesure d'organiser les sciences humaines, de remédier aux problèmes terminologiques de l'Histoire. D'ailleurs, c'est probablement la raison pour laquelle beaucoup de chercheurs en sciences humaines prônent l'organisation terminologique, bercés par ce rêve – inaccessible évidemment – d'une langue claire et limpide à la traductologie, d'une unification des termes, d'une normalisation des concepts, etc. Ils y voient le moyen d'atteindre une perfection terminologique – voir conceptuelle - qui ne peut qu'être bénéfique à leur discipline. En effet, la terminologie apparaît comme la discipline la plus apte à fixer les termes, à brider les imperfectio*ns, a* combler les lacunes. L'ordre qu'imposait une telle discipline face au désordre qui règne parmi les termes de la traductologie ne peut qu'être acclamé, voire admiré. Le rôle de la terminologie devient thérapeutique faisant presqu'écho à la « maladie de la langue » de Wittgenstein qui voyait dans la clarification des concepts philosophiques le moyen de guérir la philosophie de ses problèmes conceptuels. Or, voulant remédier à la crise terminologique auquel la traductologie faisait face, celle-ci s'est vue au milieu d'une autre crise bien plus grave : le regard négatif que la terminologie pose – indirectement - sur la traductologie. En effet, la terminologie, voie unique à prendre vers la précision conceptuelle, se devait de stabiliser la traductologie pataugeant dans une mare d'irrégularités conceptuelles. C'est comme si l'on concédait à la première le droit et la capacité à brider l'élan déréglé de la deuxième. Comme si la traductologie était, à l'instar de toutes les sciences humaines, irrationalité et incohérence. Et c'est sans doute cette idée implicitement négative qui a empêché les traducteurs et terminologues traitant des sciences humaines à aller au-delà d'une énumération évidente et superficielle des problèmes terminologiques particuliers de ces domaines ; ils multiplient arguments et exemples démontrant la singularité de la terminologie des sciences humaines, mais s'arrêtent à un constat décevant : il est d'autant plus important d'appliquer un travail terminologique à ces domaines. Même quand ils mettent le doigt sur les incohérences de la théorie classique appliquée aux sciences humaines, c'est pour davantage condamner les sciences humaines que modifier les théories de la terminologie. Leurs critiques restent donc tâtonnement et

---

18 ESCARTIN, Arilla, "Le travail terminologique pour la traduction de l'histoire », in *Mots, termes et contextes*, Actes des Septièmes Journées Scientifiques du réseau de chercheurs "Lexicologie, terminologie, traduction" Bruxelles, 8-9-10 septembre 2005, Ed. des arches contemporaines, Coll. Agence universitaire de la francophonie, Paris, 2006, p.104.

hésitation, toujours incapables de vraiment proposer une théorie alternative qui puisse être plus apte à accueillir toutes les disciplines. Si la socioterminologie aide à assouplir les présupposés théoriques classiques, elle ne résout pas pour autant les problèmes de la traductologie. Elle permet certes de nuancer la vision presque sacro sainte de l'indépendance du terme par rapport à son contexte, mais reste en deçà d'une véritable étude menée sur les sciences humaines. C'est ainsi que la terminologie appliquée aux sciences humaines peut certes bénéficier des découvertes de la socioterminologie mais ne peut s'y limiter sinon elle ne sera jamais apte à révéler la nature profonde des termes en traductologie. Celle-ci exige une analyse différente. Bien mieux, elle offre à l'analyse une occasion privilégiée de revisiter la terminologie. La socioterminologie laisse la place à une marge d'humain, mais étroite ; à une évolution, mais lente. Où sont donc ces possibilités cachées, ce potentiel dissimulé ? Le sujet est complexe, important et mérite de s'y attarder car la question des sciences humaines et de la traductologie en l'occurrence oblige à penser autrement la terminologie.

Afin que la terminologie soit plus apte à s'appliquer à la traductologie, elle doit tout d'abord changer la vision qu'elle a des sciences exactes. C'est après avoir brisé l'attachement de l'«objet» à ces sciences-là que l'on peut détacher la terminologie des domaines scientifiques. Il est donc nécessaire d'interroger le triangle sémiotique et l'objet en particulier. Et c'est ce que la socioterminologie a omis, ou presque, de faire. Contrairement à Rey qui lui critique brutalement la théorie classique, anéantissant impitoyablement ses fondements sans jamais proposer d'alternative, Depecker se livre plutôt à une pondération de cette théorie relevant souvent ses faiblesses et attirant l'attention sur les non-dits qui y demeurent. Il préfère ainsi nuancer la théorie classique plutôt que de la rejeter essayant timidement de rester dans la lignée classique tout en apportant des éléments de réponse aux lacunes persistantes. Dans un passage très significatif, il dit que « la question qui se pose alors est celle de la sélection des différents caractères retenus pour la définition. [...]. Elle [cette approche du concept] repose sur la constatation que le traitement et l'organisation des connaissances conduisent à considérer le concept comme un élément à plusieurs dimensions. Chacune de ces dimensions est déterminée par les critères retenus pour décrire le concept. Ces critères peuvent varier en fonction de plusieurs paramètres, notamment le domaine choisi, l'angle de vue considéré, la direction de la recherche envisagée, l'intérêt de dégager telle propriété de l'objet plutôt qu'une autre, etc."[19] (c'est nous qui soulignons) On voit déjà clairement qu'un tel passage n'est pas moins qu'une véritable rupture avec la théorie classique : il y a soudainement ce côté « personnel » ou « personnalisé » de la création du concept, cette attestation explicite de la participation subjective du sujet à cette création. Car, en fin de compte, tout concept n'est autre que le résultat d'un certain nombre de

19 DEPECKER, Loïc, DEPECKER, Loïc, Entre signe et concept : Eléments de terminologie générale, Ed. Presses Sorbonne Nouvelle, Paris, 2002, p.85.

choix effectués par les individus selon un certain nombre de critères que Depecker nomme « la dimension du concept »[20]. Il rajoute même que "le moindre travail terminologique en effet, s'il veut être fidèle aux objets qu'il décrit, se doit de prendre en considération les variations géographiques, temporelles, culturelles, psychologiques, voire psychanalytiques (Depecker, 1995, (2.)).[21] Cet auteur nous fait passer dans cette citation de l'univers du réel toujours revendiqué par la terminologie *a un un*ivers de possibles. Il n'y a plus « le » concept unique qui reflète les propriétés substantielles d'un objet, mais des possibilités infinies d'appréhender ce concept. L'objet est certes toujours là, mais le concept ne révèle plus ses *qualit*é*s int*rinsèques, il se construit grâce aux propriétés que le terminologue « estime » pertinentes pour son étude. Cependant, ce genre de remarques demeurent très sporadiques, incapables de se constituer en théorie générale ; elles sont souvent déduites d'une longue pratique terminologique qui voit les failles de la théorie classique, mais sont souvent énumérées l'une à la suite de l'autre sans que ne soit portée plus loin l'analyse théorique. Depecker a donc une intuition qu'il n'investit pas à fond, mais qui ouvre la porte à une remise en question de la conception ontologique de la terminologie. L'on déduit tout de même que la terminologie ne peut continuer à s'enfermer dans l'absoluité d'une terminologie de la substance. Le cœur du problème réside dans le fait qu'elle considère la science comme l'expression d'un être stable, donné dans la nature. Mais les questions qui ont pu un jour se poser relativement à l'objectivité de la connaissance, ou au déterminisme éventuellement présent dans la nature, reflètent les préoccupations de l'ancienne physique. En effet, les phénomènes physiques étaient scrutés selon une conception « chosiste » de la nature qui poursuivait ce rêve d'une science objective en calque direct sur le monde. Or, la science ne peut prétendre à une lecture transparente de l'être. La connaissance scientifique ne relève pas d'une adéquation directe à la chose, elle n'embrasse nullement les formes de l'objet dans ses moindres détails réels; bien au contraire, elle découle vraisemblablement d'une réflexion incessante sur notre expérience et d'une remise en cause continuelle de celle-ci. C'est ainsi que le progrès scientifique en général se fait, non par l'observation attentive de la réalité de la nature, mais plutôt par une remise en question perpétuelle de cette prétendue réalité de l'objet. L'ontologie d'hier se rattachait donc à des préjugés « naturalistes », une ontologie de la substance naturelle. Cependant, cette vision de vérité-adéquation où l'objet est posé dans la nature, attendant que l'homme le découvre, est une pure illusion et une source d'illusion. Reprendre cela, c'est doter nos raisonnements de faux immédiats où la compréhension se fait sans effort. Cette idée de l'existence d'une immédiateté ontologique se déduit d'une rationalité scientifique qui se positionne face à un objet qu'elle considère posée dans une immuabilité à découvrir. Cependant, l'objet scientifique, et c'est le leitmotiv de Bach*elard, es*t un

20 DEPECKER, Loïc, Ibidem, p.121.

21 DEPECKER, Loïc, Ibidem, p.77.

objet construit, jamais donné. Voilà donc que la science bachelardienne renonce à l'ontologie de la substance traditionnelle pour adopter une ontologie de la relation. En fait, parlant des problèmes à l'œuvre dans les sciences, Bachelard affirme que « les problèmes ne se posent pas d'eux-mêmes »22 parce que la nature ne nous informe jamais des questions que le scientifique devrait lui poser. Quoi qu'on en dise, il y a vraisemblablement une part d'arbitraire dans la recherche scientifique qui fait que l'on n'emprunte jamais, lors du questionnement scientifique, l'unique et seule voie d'accès à la réponse. On ne peut donc dire qu'il y a une évidence naturelle des questions scientifiques que se pose le savant. D'habitude, c'est de la confrontation avec les écrits d'autres scientifiques que sont suscités nos questions sur le monde et non selon une contemplation et une observation solitaire du scientifique sur le monde. Il n'est donc nullement vrai que le scientifique, en tant que sujet pensant et connaissant, se positionne pour ainsi dire dans une sorte de « cogito esseulé » face à un monde mystérieux à déchiffrer. Revenons un peu à la vision classique que la terminologie a de l'objet : Une entité spatio- temporelle matérielle «en attente» de conceptualisation (concept) et de dénomination (terme/signe). Cette vision présente un objet qui s'offre à l'homme dans une idéalité et une simplicité presque naïves si j'ose dire. C'est ainsi que la philosophie bachelardienne non seulement remet en cause la théorie terminologique héritée d'une tradition naturaliste ancienne, mais l'anéantit totalement. L'objet bachelardien et l'objet te*rminologique* s'excluent mutuellement si bien *que la const*ruction de l'un ne peut qu'aboutir à la destruction de l'autre. Mais que reste-t-il donc de la théorie terminologique ? L'oubli d'une ontologie de la substance n'est-elle pas préjudiciable à la pensée terminologique ? En troquant la stabilité de l'objet donné par l'instabilité de l'objet construit, ne court-on pas le risque de voir sapée toute la raison d'être de cette discipline ?

De Platon à Frege en passant par Aristote, Descartes, Saussure, Wittgenstein... une tradition de réflexion sur le concept, les Idées, l'objet, l'être, le réel, le signe... s'est développée, mettant côte à côte et face à face des schémas théoriques et des représentations ontologiques extrêmement différents. Il y a toujours eu des divergences dans la conception que les philosophes et les linguistes ont eu de ces éléments fondamentaux. C'est surtout des tensions « antinomiques » s'organisant autour de l'être (matérialisme/idéalisme) et du concept (nominalisme/réalisme) qui ont agité les débats philosophiques et ont tenté d'apporter des éléments de réponse. Les grilles d'analyse ne sont évidemment pas les mêmes, elles ne sont par ailleurs ni simples ni univoques. Les fondateurs de la théorie terminologique ont plutôt été dans une lignée d'idéalisme scientifique qui relie Platon (idéalisme) à Auguste Compte (positivisme). Or, comme nous l'avons vu avec Bachelard, cette voie n'est pas la seule et unique voie qui permet de penser l'objet, le concept et le signe. Remettre en cause l'objet tel que le perçoit la terminologie n'est donc fatal que pour cette perception singulière de l'objet, et non pour la terminologie en soi qui pourrait, elle, trouver une autre manière d'appréhender cet élément fondamental. Si c'est pour mieux refléter la nature de la traductologie que cette remise en cause a été initiée,

c'est paradoxalement dans le désillusionnement des sciences exactes, et de toutes les sciences, que cette remise en cause a été la plus fatale. C'est un peu remettre tous les domaines de la connaissance *a pied d'*egalité, car c'est bien de connaissance qu'il s'agit. Nulle science n'est capable de dire le réel, c'est toujours dans l'univers de la connaissance et du savoir, que notre réflexion s'aventure. Même dans le cas des sciences qui prétendent avoir accès au réel comme la physique, il y a toujours des révolutions qui viennent remettre en cause la valeur absolue de toute connaissance scientifique. La terminologie est donc restée longtemps emprisonnée dans cette illusion de réalité alors qu'elle évolue dans des domaines qui ne parlent que de connaissance. Si la terminologie scientifique s'en est pendant longtemps plus ou moins sortie, c'est parce que la relativité – donc subjectivité – dont elle fait en réalité preuve est quantifiable, formalisable. Pendant longtemps, ce côté empreint d'un semblant d'objectivité a permis de nourrir l'illusion du réel dans la concrétude directe ou l'abstraction axiomatique. Elle prend pour réalité une subjectivité quantifiable, ce que d'aucuns ont nommé une intersubjectivité[22], qui entretient un mirage réaliste. C'est pour cela que la terminologie a pu jusqu'à présent plus ou moins se pencher sur les sciences exactes pures et appliquées. Mais, aussitôt formé, ce mirage se dissipe rapidement dans les sciences humaines. C'est en substituant, non seulement l'objet donné par un objet construit, mais aussi l'objet réel par un objet de connaissance, que la terminologie peut enfin revendiquer une prétention à la généralisation. Cela est d'autant plus pertinent que la différence entre les termes et les mots ne réside pas tant dans la nature de l'objet, du concept et du signe, que dans leur fonction et leur statut au sein du domaine auxquels les termes appartiennent, c'est dans leur capacité à véhiculer la connaissance du domaine dans une sorte de concentration conceptuelle particulière. L'erreur fatale de la terminologie a été de confondre ces deux objets foncièrement divergents, or aucune science, comme nous l'avons vu, ne traite du réel. Partant, les concepts ne « conceptualisent » plus un objet du monde mais un objet de connaissance. L'ontologie du réel est rejetée à l'extérieur de la connaissance qui ne se préoccupe que de la construction de l'objet par le sujet connaissant. Mais quel est cet objet de connaissance ? Est- ce une simple substitution formelle à l'objet réel sans que ne soit bouleversé le caractère profond de l'objet ? Et en quoi cela nous rapproche-t-il de la traductologie ? L'objet de connaissance est en quelque sorte opposé à l'objet réel. A la transcendance, l'unifacialité de l'un s'oppose la complexité, la multiplicité de l'autre qui fait écho à l'univers subjectif de la traductologie. L'objet de connaissance permet en effet de réunir une multitude de théories sur l'être: du nominalisme, au réalisme, en passant par le conceptualisme, l'objet de connaissance absorbe en quelque sorte tous ces éléments pour la simple raison que la connaissance en soi est protéiforme. Pour le nominalisme, position classique soutenue par la terminologie, il n'y a de dénoté que des représentations d'unités spatio-temporelles. Pour le réalisme par contre, le

22 Possibilité à plusieurs personnes d'emprunter la même voie pour arriver à la connaissance.

dénoté est double, à la fois représentation des Idées et représentation des êtres spatio-temporels. Le conceptualisme enfin soutient que, comme le nominalisme, le dénoté est issu de représentations spatio-temporelles, mais que comme le réalisme, ce sont déjà des concepts qui constituent de telles représentations. Son ontologie est donc complexe et stratifiée. En somme, l'on ne trouve que deux objets réels, la chose et la conscience humaine (« je pense, donc je suis »). Par contre, en se plaçant sur le plan de la connaissance, il est des êtres concrets (les unités spatio-temporelles qui ne se confondent pas pour autant avec la chose) et des êtres conceptuels abstraits (les Idées). L'on peut même trouver des êtres issus d'un mélange de concret et d'abstrait. Il est indéniable qu'une telle conclusion nous rapproche davantage de la traductologie. Pour la première fois en terminologie, l'objet en tant qu'être purement abstrait ou psychique peut trouver un véritable point d'appui, une reconnaissance de son existence ontologique. C'est donner enfin une légitimité à une « réalité » non issue de la concrétude pure du monde, à une ontologie où la conscience humaine joue un rôle majeur. Voilà les potentiels cachés de la terminologie. C'est en quelque sorte faire marche arrière et redonner de la légitimité à toute la dimension subjective instable de la traductologie qui fait désormais partie intégrante du processus de connaissance. Que la déverbalisation n'est pas d'ancrage concret dans la réalité, qu'elle soit incapable d'être formulée par des lois ou des formules mathématiques ne lui enlèvent en rien son droit à être « terminologisée ». C'est donc

en changeant le regard que la terminologie porte sur son propre objet, sur son « être » à elle, qu'elle est enfin capable de légitimer en quelque sorte la nature véritablement mouvante de la traductologie. La terminologie limite alors sa tendance normative pour être plus à l'écoute de disciplines telles la traductologie ; elle est plus dans le mouvement naturel de la connaissance que dans la rigidité for*melle d*es codes. Il n'est plus un modèle idéal d'«objet » tout-puissant à admirer, mais un objet qui avance avec le mouvement du domaine concerné. L'objet de la connaissance est donc diversiforme, construit. Et c'est parce qu'il est construit, qu'il est en construction permanente. Toute science, est un esprit spéculatif en progression, sans cesse rectification et reprise. Ce n'est pas le définitif qui constitue sa marque mais la volonté permanente d'y accéder. Comme nous l'avons dit, la science contemporaine est en mesure de transformer ses propres objets et l'être n'est pas une fonction absolue traversée par un seul et unique paradigme ra*tionnel. Elle est relec*ture permanente des acquis, *mais les relectures* sont différentes : les sciences exactes effacent et remplacent tandis que la traductologie reprend et rajoute … La pensée scientifique est sans arrêt gommée, changée, remplacée parce qu'elle supprime les anciennes découvertes tombées en désuétude offrant donc à la rationalité les conditions de se construire en se spécialisant avec le temps. Les recherches en traductologie tiennent davantage à conserver les thèses passées et à les revisiter, chaque penseur ayant plutôt tendance à affermir ou rajouter qu'à supprimer et *remplacer*. C'est probablement la raison pour laquelle il ne peut y

avoir de circonscription de l'objet traductologique, juste une tentative de circonscription. A ce titre, la transformation incessante des termes et des concepts apparaît inévitable.

A Formal Ontology for Industrial Maintenance

Mohamed Hedi KARRAY*, Brigitte CHEBEL-MORELLO*, Noureddine ZERHOUNI*

* Automatic Control and Micro-Mechatronic Systems Department, 24, Rue Alain Savary, 25000 Besançon, France
Hedi.karray@femto-st.fr
Brigitte.morello@femto-st.fr
Noureddine.zerhouni@femto-st.fr
http://www.femto-st.fr/fr/Departements-de-recherche/AS2M/Presentation/

Abstract. The rapid advancement of information and communication technologies has resulted in a variety of maintenance support systems and tools covering all sub-domains of maintenance. Most of these systems are based on different models that are sometimes redundant or incoherent and always heterogeneous. In order to create a common and shared language, we are developing a domain ontology of industrial maintenance adopting METHONTOLOGY approach to manage the lifecycle development of this ontology that we have called IMAMO (Industrial MAintenance Management Ontology). The ontology developed can be exploited to ensure semantic interoperability and to generate new knowledge that supports decision making in the maintenance process.

1. Introduction

Industrial maintenance is a fundamental function in the business process and thus requires the development of computer systems [1]. It covers all domains of a business, from the plant and the equipment to be maintained, to the organization according to different strategies (preventive maintenance, predictive maintenance, corrective maintenance), to managing operators and material (handling, hoisting) and spare parts, to the computer-assisted diagnostic systems, to documentation management, etc. Various maintenance support systems and tools have begun to involve and to be essential for maintenance process management to cover all these sub-domains as CMMS, diagnostic support system, prognostic system, resource management systems as (ERP) and other systems. All these systems are currently based on different models that are usually complementary, but sometimes redundant, sometimes incoherent and always heterogeneous.

The principal problem confronted by these systems is to provide the means to move from coexistence to the interoperability and cooperation of different applications within the same environment.

Indeed, different researchers [2,3,4 and 5] are in agreement that Ontology engineering is recognized as the key technology to deal with this. Ontologies specify the semantics of terminology systems in a well defined and unambiguous manner [6], by formally and explicitly representing shared understanding about domain concepts and relationships between concepts. With Ontology intended meanings of terminologies and logical properties of relations are specified through ontological definitions and axioms in a formal language. The employment of a consistent Ontology is useful because maintenance support systems' users can have identical semantics for all concepts. All maintenance systems share the common Ontology of maintenance, so have an interoperable communication.

Additionally, we consider that a common neutral model is the most feasible solution in maintenance systems to enable data sharing or integration in heterogeneous applications. With a common Ontology model, it is possible for maintenance information to be created once, re-used and enriched during the equipment lifecycle. This reduces maintenance duration by eliminating the need to recreate data models repeatedly and increases maintenance quality by eliminating errors and using experience of previous maintenance activities. Ontology model organizes data and knowledge of a certain domain of interest (application) in a manageable manner. It should contain the definitions of all application objects (e.g. diagnostic, documentation) within that domain, as well as constraints and relationships between objects.

This, sharing of knowledge via the ontology allows to each system to exploit all the knowledge of the other systems. As well, the reasoning methods which can be applied on the ontology provide an added value to this knowledge while it can generate new knowledge that users cannot notice. This generated knowledge can be

exploited by all systems and then gives a surplus to maintenance actors usually in the aim to provide the right information in the right format to the Right people to do the Right things at the right time.

For these reasons, we provide in this work (in the scope of SMAC project[1]) a domain ontology of maintenance domain to be shared and integrated in the maintenance platform as well as its integrated applications.

When a new ontology is going to be built, several basic questions arise related to the methodologies, tools and languages to be used in its development process [5]. From several methodologies developed to formalize ontologies creation we adopt METHONTOLOGY approach [9].

In the formalization phase, a UML ontological model is built in collaboration with maintenance experts; this ontology encloses 8 models classified as unavoidable for the maintenance process which are (1) the structural model presenting the equipment composition, (2) the functional model which characterizes different functionalities of the equipment and its components, (3) the dysfunctional model, (4) the event model which defines the triggering event launched after failures and/or degradation, and (5) the informational model which presents various resources (documents, human, software, tools…) related to the equipment and maintenance tasks, (6) the interventional model presenting concepts related to the intervention process, (7) teh Strategy model presenting managerial aspects about maintenance strategy and contracts and (8) Process model presenting all technical, administrative and managerial process.

After that, to make the ontology operational, in the implementation phase of METHONTOLOGY we transform the UML model to PowerLOOM, a logic-based representation language for ontology presentation. Some proposed metrics [8] are then used to characterize the ontological model.

The remainder of this paper focuses the development process of IMAMO (Industrial MAintenance Management Onotlogy). Hence the rest of sections will be organized as follows. Section two is devoted to addresses the first three phases specification of METHONOLTOGY which are knowledge acquisition and conceptualization of the ontology. Section three is dedicated at the development of the formalization reusing and implementation of IMAMO. In section 3, the evaluation phase, a business-oriented approach based on cases of use is adopted to evaluate the ontology at the application level. Finally, we conclude the paper and we give some perspectives and guidelines to future work.

[1] http://smac.univ-fcomte.fr

2. IMAMO development process

Fernandez and al. in [9] assert that the ontology development process refers to the activities needed in order to build ontologies. In this context, METHONTOLGY has been proposed as a structured method to build ontologies. This methodology includes a set of activities, techniques, and deliverables to be produced after the execution of such activities using the corresponding techniques. The set of activities is {planning, specifying, acquiring knowledge, conceptualizing, formalizing, integrating, implementing, evaluating, producing documents, and maintenance}.

2.1 Specification

Domain	Industrial maintenance
Name	IMAMO: Industrial MAintenance Management Ontology
Date	2010
Conceptualized-by	Mohamed-Hedi Karray, Brigitte Morello, Thibault Bobyck
Implemented-by	Mohamed-Hedi Karray, Thibault Bobyck
Purpose	Ontology concerns most concepts of industrial maintenance when information about all technical, administrative and managerial activities and actions is required in maintenance information systems. This ontology can be used to ascertain decision making throughout the life cycle of maintenance activities from failure detection to intervention and repair.
Level of Formality	Formal
Scope	Structure of equipment to be maintained, spare parts, monitoring activity, failure detection, events, material resources, maintenance actors, technical documents, administrative documents, intervention, maintenance reports, equipment states, equipment life cycle....
Sources of Knowledge	Standards (AFNOR, MIMOSA..), projects, experts

TAB. 1 – *Ontology requirement specification document*

2.2 Knowledge acquisation

Creating an ontology is not a trivial task. It requires not only the skills of information technologies but also a significant knowledge in the modeled domain [10]. Hence, as mentioned in the specification document, we based our knowledge acquisition on standards, research projects and industrial experts in maintenance. Concerning standards, we adhere to the AFNOR norms and MIMOSA standards. The PROTEUS and PROMISE projects also serve as a base. Finally, we adopt the business expertise of various maintenance experts, managers and operators from different companies such as Cegelec SA France & Germany, Tornos (Switzerland), Peugeot (Belfort, France) and em@systec (France). Various research works such as [11, 12 and 13] are also taken into account.

AFNOR defines maintenance as “the combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function”. In concordance with this definition, Retour et al. [12] present the maintenance function as a set of activities grouped into two subsets: activities with technical predominance and activities with management predominance.

Consequently, the concepts which should be identified must cover all these layers and activities. To facilitate our identification of concepts, we identified all those concepts related to each layer presented above. Identification of the main concepts of each layer is based on the models of MIMOSA CRIS, the PROTEUS project and the PROMISE SOM and SMAC projects.

2.3 Conceptualization

In this phase, Gomer-Perez et al recommend a set of intermediate representations for conceptualizing a domain ontology of objects [9].

2.3.1 Glossary of terms and data dictionary

Hence, we begin the conceptualization of IMAMO by building the glossary of terms. The concepts are first classified in the glossary respecting the four layers identified by Rasovska et al. presented above. We then refined this list of concepts by deepening the first classification. The second classification goes further than the first by breaking each layer down into sub layers. We note that some concepts are redundant in different layers or sub-layers. This is done purposely in order to count all the concepts in each layer so as to obtain a clearer vision and a more precise identification. Next we edit the data dictionary based on the European norm NF EN 13306:2001 published by AFNOR. Due to limits of space we will not show these steps in three separate tables, but we will present only the data dictionary in the formalization phase in order to gain a better understanding of the different models of the ontology (see next sections).

We note, however, that IMAMO will be a generic ontology; different details can be neglected and left to users according to their needs. In this case, users may adapt, involve and maintain the ontology.

2.3.2 Concept classification trees

After this classification, we edit the concept classification trees. We notice that the domain is very broad; nevertheless, the ontology that we develop will not contain a lot of trees. This is due to the aim of obtaining a rich ontology with different types of relations and not a hierarchical ontology like taxonomy. *Is-a, is-component-of, has* and other verbs are the relations supported by the ontology. Figure 1 summarizes concept classification trees of resources in IMAMO (i.e. is-a relations).

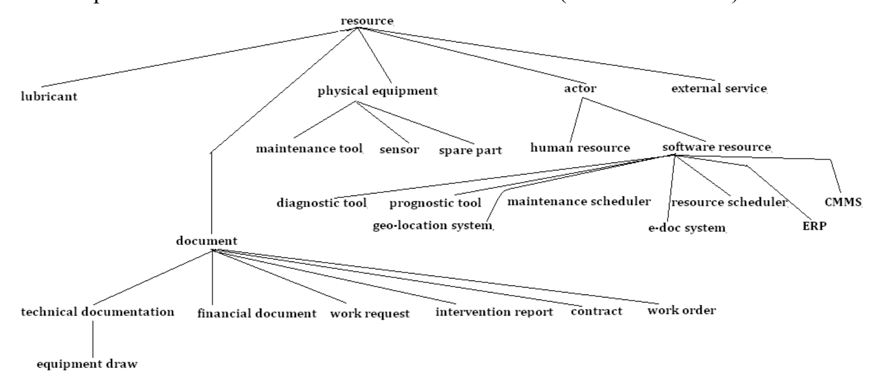


Fig. 1 – *classification trees of resource in IMAMO*

2.3.3 Edit rules

As mentioned above, rules will be left to user choice. We give here just some examples of rules that can be edited. We edit these rules by the description logics ALCQHI. Then rules can be translated and edited by the implementation language if it allows rules definition. Rules enrich the Ontology and allow more semantic reasoning as well as understandability.

For example thanks to the following defined rule the identification of critical components is possible without defining a new concept called critical component. The rule describe critical component as any physical equipment having the value of the propriety Functional-degree greats or equals to five.

$$\text{CriticalComponent} \equiv \text{PhysicalEquipment} \sqcap (\leq _{FUNCTIONNAL-DEGREE}\ 5)$$

3. Formalization re-use and implementation phases

3.1 Formalization

The Unified Modeling Language (UML) would be a good candidate for representing ontologies and knowledge [14]. Knowledge expressed via UML is easily accessible for human comprehension. In an UML model, Knowledge can be changed easily due to the modular nature of object-oriented modeling. Also, new knowledge can be derived from UML models by reasoning about their contents [14]. With this standpoint, UML can be regarded as an appropriate candidate for knowledge representation. In this work we adopt the UML class diagram to formalize IMAMO. This choice is argued by the graphical expressivity and the semantics power of UML recommended in the various research works mentioned above. This fact facilitates the exchange between domain experts and the human understanding of the ontology. In the other hand, the ontology of the domain, although formalized independently of the methods of reasoning has a structure which depends on how acquired knowledge will be used for reasoning because experts deliver the knowledge adapted to their reasoning. Reasoning methods will be considered in the implementation phase.

We decompose the class diagram into seven models considered as inescapable for maintenance process according to the layers classification used in the conceptualization phase. These models are:

(1) The structural model presenting the equipment composition; related to the equipment analyses layer.

(2) The event model which define the triggering event launched after failures and/or degradation; related to the fault diagnosis and expertise layer.

(3) The functional and dysfunctional model which characterizes different functionalities of the equipments and its components as well as which is related to the fault diagnosis and expertise layer.

(4) The informational model which presents various resources (documents, human, software, tools, indicators…) related to the equipment and maintenance tasks as well as maintenance strategy and processes; related to resources management and maintenance strategy management layers.

(5) The interventional model presenting concepts related to the intervention process.

(6) The strategy model presenting managerial aspects about maintenance strategy and contracts.

(7) The process model presenting all technical, administrative and managerial process.

Below we present the tow first models with the data dictionary containing definitions of the modeled concepts.

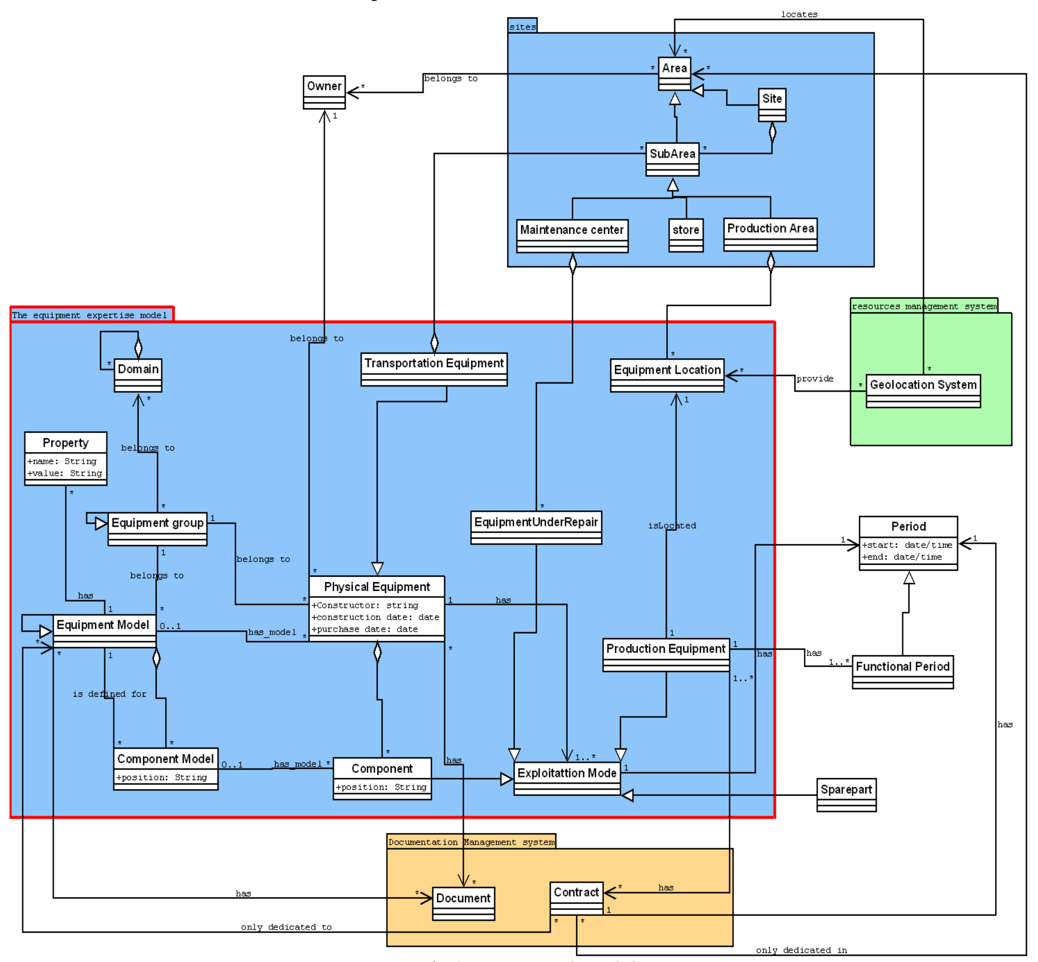


Fig 2– *Structural model*

Concept Name	Synonyms	Description
domain;		A particular field of knowledge or expertise (e.g. hydraulic).

physical equipment;	Asset Physical- product Machine Device Item	Tangible, instantiated, serialized object, component, device, subsystem, functional unit, equipment or system that can be individually considered to be maintained. A physical equipment may be an entire facility, an entire functioning platform (such as an CH-47 Tail Number XYZ helicopter), or a component piece of equipment, such as a specific instance of a bearing.
transportation equipment;		Particular Physical Equipment conveyance. A conveyance which may contain one or more area (s) of production, a set of maintenance teams, and a set of stores. For example: A fishing vessel off the coast has its own decomposition (motor allowing it to move...) as well a production area that cleans the fish and freeze.
maintenance tool;		Particular physical equipment that used as tool for performing the maintenance activities. This type of physical equipment has to be maintained also.
equipment model;	As-designed-product Model	Conceptual view of physical composition of the equipment. It is composed by the various components models of the components composing a physical equipment.
component model;		Conceptual view of a component (e.g. model of electrical motor).
Component mode;		Component is an exploitation mode that can be played by a Physical Equipment. It has the particularity to be within superior physical equipment (e.g. motor3X57H).
exploitation Mode;		Abstraction of a role played by equipment. It presents the state of exploitation that can take a physical equipment. It can be exploited as a component, a production equipment, a spare part or be under repair.
equipment under repair mode;		Specific exploitation mode affected to a Physical Equipment while it is during repair or located in a maintenance center to be repaired.
production equipment mode;		Specific exploitation mode affected to a Physical Equipment while it is exploited in a production tasks and/or located in a production area.
Spare part mode;		Specific exploitation mode affected to Physical equipment intended to replace corresponding physical equipment in order to restore the original required function of the physical equipment. Generally, it is located in a store.
equipment location;		Position of a physical equipment in a production area (to locate and track the positions of equipment).
area;		Particular geographical region (to manage multisite).
sub area;		Region that makes up part of an area.
site;		Place or setting of something. An area or plot of ground with defined limits on which a building, project, park, etc., is located or proposed to be located.
maintenance	Maintenance	Specific area to do maintenance tasks.

center;	workshop	
store;		Stock or supply reserved of physical equipment for future use.
production area;		Specific area to do production tasks.
period;		Time interval.
functional period;		Typical period during which the equipment must perform certain functions.

Tab 2 – *Data dictinary of the structural model*

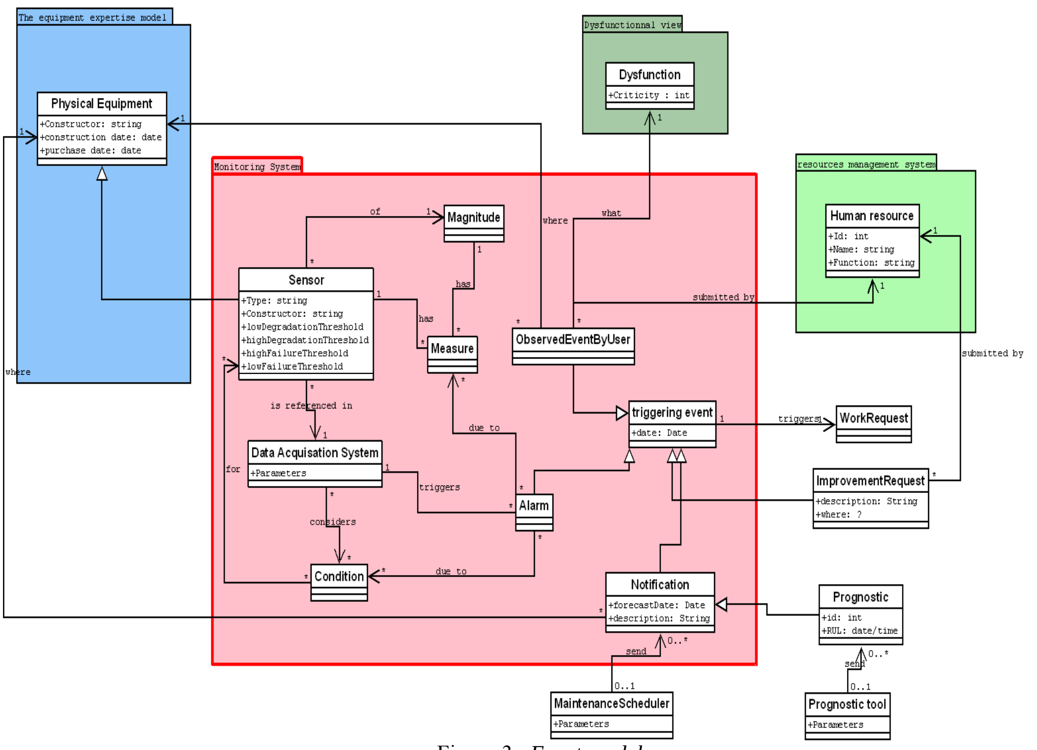


Figure 3– *Event model*

Concept Name	Synonyms	Description
measure;	measurement	Number or measure or quantity captured by a sensor.
magnitude;		Greatness of size or amount. It presents the property of relative measure.
Data acquisition system;		Software system (abbreviated with the acronym DAS or DAQ) typically convert analog waveforms

		generally retrieved from sensors into digital values for processing.
condition;		Environmental or functional requirement defined to supervise (monitoring task) a specific physical equipment or a place (e.g. site) by the use of sensors and data acquisition systems.
triggering event;		Something that happens at a given physical equipment and time that triggers a specific process of maintenance which is work request process.
alarm;		Type of triggering event launched from a data acquisition system indicating that there is a measure from a sensor violating some conditions about a specific equipment or environment.
improvement request;		Triggering event about a specific or general asking for the improvement of a physical equipment. Knowing that an improvement is the combination of all technical, administrative and managerial actions, intended to ameliorate the dependability of a physical equipment, without changing its required function.
event observed by user;		Type of triggering event about a dysfunction on a physical equipment observed by user which is a human resource.
notification;		Type of triggering event informing about future events as planned maintenance or the prognostic RUL.
prognostic;		Type of notification composed by the health status at a future time and the remaining useful life (RUL) of a physical equipment. It is the output of the prognostic tool.
Prognostic tool;		Software too or system allowing to predict and estimate the rest of time to failure and the risk of subsequent existence of one or more failure modes with a confidence level which is a value indicating the degree of certitude that the prognosis is correct.
Maintenance scheduler;		Type of software resource that allows planning, allocation of significant amount of time, and high degree of coordination between different departments, and is typically initiated through a work order. It is considered as a software resource.

Tab3– *Data dictionary of event model*

3.2 Reusing: integration and merge

While ontologies are built to be reused, Ontology reuse is one of the important issues in ontology construction. According to Pinto et al there are two different reuse

processes [15]: merge and integration. Both of these reuse processes are included in the overall process of ontology building. Merge is defined as the process of building ontology in one subject reusing two or more different ontologies on that subject [15]. Otherwise, in an integration process source ontologies are aggregated, combined, assembled together, to form the resulting ontology, possibly after reused ontologies have suffered some changes, such as, extension, specialization or adaptation. We should note that both processes are adopted in this work.

To create IMAMO we started from models developed in PROTEUS project, we involved a first version of maintenance ontology published in [16]. This ontology, is composed by 62 concepts and 70 relations integrating the main concepts used in PROTEUS. Then, in the scope of SMAC project [17] we have done a mapping between this ontology with the PROMISE model. As result of this work, Matsokitis and Karray [18] proposed an involved version of this ontology by orienting it to maintenance field when integrating some concepts related to the maintenance area included in the MOF phase of PLM. This ontology called SMAC-model is formalized by UML and implemented with OWL-dl via PROTÉGÉ tool.

Then, we come back to maintenance scope, we integrating some concepts from SMAC-model related to the life cycle of equipment to take into account the beginning of life part concerning the design phase as well as the middle of life phase by tracking all the events and health states of the equipment and the end of life by the calculation of indicators supporting decision for reuse and disassembly. An example of some reused concepts integrated in IMAMO is presented in the next table. In addition, as mentioned above, MIMOSA-CRIS is considered as the reference of the domain, for this reason, when creating IMAMO we taken account classes used in this model. A mapping between MIMOSA-CRIS and IMAMO is done, some labels of reused concepts are changed but the alignment between these concepts is added as equivalence or subsumption rules after the mapping between the two ontologies.

IMAMO Concepts	**From which ontology?**
Physical equipment	PROMISE: *Product* -- MIMOSA-CRIS: *Asset*
Equipment Model	MIMOSA-CRIS: *Model* -- PROMISE: *As-designed-product*
Equipment group	MIMOSA-CRIS: *Asset type* -- PROMISE: *product group*
Site	MIMOSA-CRIS: *site* – SMAC-Model: *Location site*
Triggering event	MIMOSA-CRIS: *Event type* -- PROMISE: *Event*
Measure	MIMOSA-CRIS: *Measurement Event* -- PROMISE: *Field Data*
Equipment location / Geo-location system	MIMOSA-CRIS: *Geoposition*
Alarm	MIMOSA-CRIS: *Alarm type* – SMAC-Model: *Alarm*
Process pattern	MIMOSA-CRIS: *Work Management Type* -- SMAC-model: *process*

Table 10- Example of reuse table

In the other hand, other integration is possible as of the functional ontology proposed by Mizoguchi et al [19]. This ontology can be integrated under the concept Function The ontology of faults presented by Kitamura et al [20] can be integrated to involve the dysfunctional model of IMAMO. As well as, task ontologies as the one presented in [21] by Ikeda et al can be integrated under the concept Task.

3.3 Implementation

We translate the UML model of IMAMO to PowerLoom. We chose to work with PowerLoom version 3.2.0 despite version 4.0 is actually available because the version 3.2.0 is a stable version unlike to the forth version which is a beta version yet.

In addition, we note that a PowerLoom exporter for the Protégé editor has been implemented. It can write ontologies using the Protégé frame language in PowerLoom, either fully native or with support for the system concepts from Protégé. Moreover The PowerLoom GUI (or knowledge editor) a Java-based graphical client for PowerLoom is now standardly included and available with PowerLoom starting with version 4.0 [22].

We present in the following a part of the structural model of the equipment implemented by PowerLOOM. Each UML class is translated into a PowerLOOM concept using ″DEFCONCEPT″ command. Associations and attributes of classes are translated into PowerLOOM relation or function using ″DEFFUNCTION″ and ″DEFRELATION″ commands.

```
(DEFMODULE "/PL-KERNEL-KB/PL-USER/ONTOLOGIE-MAINTENANCE"
  :DOCUMENTATION "Module for Maintenance"
  :INCLUDES ("PL-USER"))

(IN-MODULE "/PL-KERNEL-KB/PL-USER/ONTOLOGIE-MAINTENANCE")
(IN-DIALECT :KIF)
(DEFCONCEPT COMPONENT)
(DEFCONCEPT PHYSICAL-EQUIPMENT)
(DEFRELATION PHYSICAL-EQUIPMENT-ID ((?C PHYSICAL-EQUIPMENT)
(?ID STRING)))
(DEFRELATION PHYSICAL-EQUIPMENT-TYPE ((?C PHYSICAL-EQUIPMENT)
(?TYPE STRING)))
……….
(DEFCONCEPT EQUIPMENT-MODEL)
(DEFRELATION EQUIPMENT-HAS-TOP-MODEL ((?E PHYSICAL-EQUIPMENT)
(?MG EQUIPMENT-MODEL)))
(DEFRELATION EQUIPMENT-MODEL-INHERITS ((?MG1 EQUIPMENT-MODEL)
(?MG2 EQUIPMENT-MODEL)))
(DEFRELATION EQUIPMENT-COMPONENT-COMPOSED ((?E PHYSICAL-
EQUIPMENT) (?C COMPONENT)))
```

```
(DEFRELATION EQUIPMENT-COMPOSED ((?E PHYSICAL-EQUIPMENT) (?COM
PHYSICAL-EQUIPMENT)))
```

4. Evaluation and maintenance

The aim of our evaluation is to highlight the added value of IMAMO to maintenance systems and actors as well as in order to provide new users with sufficient information to promote the exploit of this ontology (the extent of the maintenance domain coverage). Another aim of this evaluation is to focus the weak points of this ontology to facilitate maintenance and evolution tasks.

Our approach of evaluation is composed on two main steps [23]: the quality of the formalization model and the business oriented evaluation containing the evaluation of applicability, efficiency and simplicity for interoperability.

4.1 Quality of the model

In accordance with Tartir et al [24], assessing the quality of an Ontology is important for several reasons including allowing the Ontology developer to automatically recognize areas that might need more work, allowing the Ontology user to know what parts of the Ontology might cause problems. Different dimensions are available to assess the quality of an ontology. We are interesting to quality metrics presented by [24]. We use metrics of schema evaluation to evaluate the success of the Ontology UML model of real-world domain of maintenance. How classes are organized, the depth, the richness, breadth, and height balance of the Ontology schema inheritance tree can play a role in a quality assessment.

To understand used metrics and bellow discussion please refers to [24]. We notice that the UML class diagram of IMAMO contains 187 Relations (P), 103 Concepts (classes) (C), 60 Subclasses (SC) and 40 Attributes (att).

The first metric is called Relationship Richness (RR). This metric reflects the diversity of relations and placement of relations in the Ontology. The RR of our ontological model is 75%. It exceeds largely the average. This means that our Ontology is not a hierarchic Ontology. It is not just a hierarchy of subclasses but it maintains a balance between inheritance relations and role associations. This is due to the inclusion of domain concepts and proves that the ontological model is business oriented and responds to business needs of maintenance.

The second used metric is called Attribute Richness (AR). The number of attributes that are defined for each class can indicate both the quality of Ontology design and the amount of information pertaining to instance data. In general we assume that the more slots that are defined the more knowledge the Ontology conveys. The AR of IMAMO is 0.38. This result shows the poverty of the ontological model in terms of attributes. The result gives as average 0.38 attributes per concept which is very low. This is known to us from the beginning because when constructing the

model we focus our interest on the concepts presentation and not its details information by the identification of attributes. Although we know that these details are very interesting for the quality and richness of the Ontology, this is a very difficult to implement because the Ontology must use expressive and unambiguous terms. In this context, it is a difficult task to be generic more than in the context of concepts. However, in our future work, collaborations with business experts are envisaged to identify attributes that may be generic for related maintenance concepts (e.g. details of classes of the UML ontological model).

The third used metric is called Inheritance Richness (IR). This measure describes the distribution of information across different levels of the Ontology's inheritance tree or the fan-out of parent classes. This is a good indication of how well knowledge is grouped into different categories and subcategories in the Ontology. This measure can distinguish a horizontal Ontology from a vertical Ontology or an Ontology with different levels of specialization. The IR obtained of IMAMO is 0.58. This result is near to the average. This shows that in the context of knowledge details, our Ontology maintains the balance between generality and explicitness. The ontological model is hybrid; it is neither vertical nor horizontal. We consider this result as a reached target, because our first goal was to build a generic ontological model to the maintenance domain, but at the same time not very light, so that it covers as possible all maintenance aspects (concepts).

4.2 Business evaluation of the ontology

4.2.1 Application use case

The application and tests IMAMO was done on the pallet transfer system. It represents a flexible production system. It is composed of five robotized working stations which are served by a transfer system of pallets organized into double rings (internal and external). Each station is equipped with pneumatic actuators (pushers, pullers and indexers) and electric actuators (stopper) as well as a certain number of inductive sensors (proximity sensors). An inductive read/write module allows to identify and locate each pallet and to provide information relative to required operation in a concrete station. The working station is situated on the external ring and contains pneumatic and electric actuators (puller, pusher, indexer, and stopper) as well as inductive sensors.

The following PowerLOOM code specifies how we can assert the `PHYSICAL-EQUIPMENT` SISTRE, the `EQUIPMENT-MODEL` PLATEFORME, as well as associations `PHYSICAL-EQUIPMENT-CONSTRUCTOR` "Bosch" of SISTRE, the top-model `EQUIPMENT-HAS-TOP-MODEL` of SISTRE which is PLATEFORME and the association `EQUIPMENT-COMPONENT-COMPOSED` to describe the composition of the physical equipment SISTRE:

```
(ASSERT (PHYSICAL-EQUIPMENT SISTRE))
(ASSERT (PHYSICAL-EQUIPMENT-CONSTRUCTOR SISTRE "Bosch"))
(ASSERT (EQUIPMENT-MODEL PLATEFORME))
(ASSERT (EQUIPMENT-HAS-TOP-MODEL SISTRE PLATEFORME))
(ASSERT (EQUIPMENT-COMPOSED SISTRE ROBOT))
 ……..
(ASSERT (EQUIPMENT-COMPONENT-COMPOSED CONVOYEUR ENTRETOISE))
(ASSERT (EQUIPMENT-COMPOSED CONVOYEUR COURROIE))
```

The java API of PowerLOOM gives the possibility to query the knowledge base (e.g. assertion of the Ontology). In this example we ask PowerLOOM engine to list all Physical equipment. The given answer is not just the physical equipment SISTRE but its composition. In our Ontology we defined that a Physical-equipment can be composed by Physical-equipments.

```
PL-USER |= (load "ontologie-maintenance.plm")
PL-USER |= (in-module "ONTOLOGIE-MAINTENANCE")
ONTOLOGIE-MAINTENANCE |= (RETRIEVE ALL (PHYSICAL-EQUIPMENT
?PE))

There are 28 solutions:
  #1: ?PE=COURROIE
  #2: ?PE=CONVOYEUR
  #3: ?PE=DETECTEUR
  #4: ?PE=ACTIONNEUR
  #5: ?PE=TEST
  …………
  #28:?PE=TAP-INT
```

The second example show how PowerLOOM reasoning engine take account inheritance (subclasses). We launch a query to list all ACTIONNEUR and then we launch a query to ask only ACTIONNEUR-ELECTRIQUE. The first answer includes the answer of the second query due the sub-class relation.

```
ONTOLOGIE-MAINTENANCE |= (RETRIEVE ALL (ACTIONNEUR ?A))
There are 8 solutions:

#1: ?A=INDEXEUR
  #2: ?A=TIREUR
  #3: ?A=POUSSEUR
  #4: ?A=S5
#5: ?A=S4
#6: ?A=S3
#7: ?A=S2
#8: ?A=S1

ONTOLOGIE-MAINTENANCE |=
(RETRIEVE ALL (ACTIONNEUR-ELECTRIQUE ?AE))
There are 5 solutions:

#1:?AE=S5
#2:?AE=S4
#3:?AE=S3
#4:?AE=S2
#5:?AE=S1
```

We provide a second example concerning two maintenance systems S1 and S2. The first uses the MIMOSA-CRIS model, and the second uses IMAMO. The equipment SISTRE fails. S1 sends an intervention request to S2. S2 needs all information about the equipment composition and previous intervention on the equipment. An equipment is modeled by the concept ASSET in the MIMOSA-CRIS model. In IMAMO an equipment is defined by the concept PHYSICAL-EQUIPMENT. We show a sequence of PowerLOOM examples to demonstrate how the reasoning engine can solve this problem and how it manages to get the information expected (needed) by S2.

Firstly we ask the PowerLOOM system of the Ontology used in S2 to list all equipment which it knows, after that if it knows that SISTRE is an equipment:

```
ONTOLOGIE-MAINTENANCE |= (RETRIEVE ALL (PHYSICAL-EQUIPMENT
?PE))
There are 3 solutions:              #2: ?PE=TEST01
  #1: ?PE=TEST02                    #3: ?PE=TEST

ONTOLOGIE-MAINTENANCE |= (ASK (PHYSICAL-EQUIPMENT SISTRE))
UNKNOWN
ONTOLOGIE-MAINTENANCE |= (ASK (not (PHYSICAL-EQUIPMENT
SISTRE)))
UNKNOWN
```

So the system of S2 doesn't know if SISTRE is an equipment. To ensure an understandable exchange between S1 and S2, it suffices to add in the knowledge base an equivalence rule defining ASSET as a concept equivalent to PHYSICAL-EQUIPMENT. This will allow to S2 to get all information about all concepts having as type ASSET.

```
(DEFCONCEPT PHYSICAL-EQUIPMENT ((?A ASSET)))
(DEFCONCEPT ASSET ((?PE PHYSICAL-EQUIPMENT)))
```

After that we assert that SISTRE is an ASSET

```
(ASSERT (ASSET SISTRE))
```

When we ask if ASSET is physical equipment the reasoning engine answer true:

```
ONTOLOGIE-MAINTENANCE |= (ASK (PHYSICAL-EQUIPMENT SISTRE))
TRUE
```

So, when the S2 system considers that ASSET as PHYSICAL-EQUIPMENT, it can obtain composition information of SISTRE from S1 while the two systems shares the same knowledge base.

```
ONTOLOGIE-MAINTENANCE  |=  (retrieve  all  (EQUIPMENT-COMPOSED
SISTRE ?x))
There are 4 solutions:

#1: ?X=ROBOT                            #3: ?X=DETECTEUR
#2: ?X=ACTIONNEUR                       #4: ?X=CONVOYEUR
```

5. Conclusion

Maintenance domain contains a variety of maintenance support systems and tools covering all its sub-domains. Contrariwise the largest part of these systems are based on different models that are sometimes incoherent and always heterogeneous. To respond this problem, we propose to create a domain ontology for industrial maintenance that we called IMAMO (Industrial MAintenance Management Ontology).

In this paper we presented the development lifecycle of IMAMO. From various exiting methodology of ontology construction, we adopted METHONTLOGY approach to manage the lifecycle of our maintenance ontology.

To develop this ontology, we are based on different standards and previous projects elaborated in the domain of industrial maintenance. The formalization of the ontology is done via UML class diagram which provides a rich conceptualization of the complex domain of maintenance accompanied by the data dictionary. To be operational, a logic-based representation language called PowerLOOM characterised by its power and efficient reasoning engine is used to implement IMAMO. Afterwards, different approaches are applied to evaluate the proposed ontology in two levels. The first one is to characterize the ontological model and the functional quality of the ontology. The second level is applied in order to show how IMAMO can resolve semantic interoperability problem.

Concerning shortage of this work needed to be addressed in future work, more collaboration with maintenance experts from industrial and academic communities to evolve and enrich the proposed ontology is needed. In addition, as perspectives, we aim set up of a web site containing documentation about the ontology and allowing the possibility to make maintenance and evolution activities.

Acknowledgment

This work was carried out and funded in the framework of SMAC project (Semantic-maintenance and life cycle), supported by Interreg IV programme between France and Switzerland.

References

1. Liyanage, J.P. and Kumar, U., Towards a value-based view on operations and maintenance performance management, Journal of Quality in Maintenance Engineering, Vol. 9, pp. 333–350
2. Heiler S. (1995) Semantic Interoperability. ACM Computing Surveys (CSUR).
3. Riichiro Mizoguchi (2004). "Tutorial on ontological engineering". In: New Generation Computing. Ohmsha & Springer-Verlag, 22(2):198-220.
4. Yang Q Z & Zhang Y. (2006) Semantic interoperability in building design: Methods and tools. Journal of Computer-Aided Design 38 pp 1099–1112. Amsterdam: Elsevier Science Publishers.
5. Obrst, L. (2003). Ontologies for Semantically Interoperable Systems. In CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management, pages 366–369, New York, NY, USA. ACM Press.
6. Guarino N. (1998) Formal Ontology and information systems. Formal Ontology and information systems. IOS Press.
7. O. Corcho, M. Fernandez-Lopez, and A. Gomez-Perez, Methodologies, tools and languages for building ontologies: Where is their meeting point?, Data & Knowledge Engineering, Vol. 46, 2003, 41-64.
8. Tatir, Samir; Arpinar, I. Budak. Ontology Evaluation and Ranking using OntoQA. Int. Conf. on Semantic Computing(ICSC), 2007
9. Gómez-Pérez, A.; Fernández, M.; de Vicente, A. (1996). Towards a Method to Conceptualize Domain Ontologies. Workshop on Ontological Engineering. ECAI'96. Budapest. Hungary. PP: 41-52.
10. FRANKOVIČ, Baltazár - BUDINSKÁ, Ivana. The role of ontology in building of knowledge systems for industrial applications. In 4th Slovakian - Hungarian Joint Symposium on Applied Machine Intelligence : proceedings. ISBN 963-7154-44-2. s. 15-25.
11. Kaffel H. (2001) La maintenance distribuée: concept, évaluation et mise en oeuvre. Phd thesis, Université Laval, Quebec.
12. Retour D., Bouche M. et Plauchu V., Où va la maintenance industrielle, Problèmes Économiques, No. 2.159, pp. 7-13, 24 Janvier 1990.
13. Rasovska I, Chebel-Morello B & Zerhouni N. (2004) A conceptual model of maintenance process in unified modeling language. Proceedings at 11 th IFAC Symposium on Information Control Problems in Manufacturing 2004 (INCOM)
14. Stephen Cranefield, Networked Knowledge Representation and Exchange using UML and RDF, Journal of Digital Information, Vol 1, No 8 (2001).
15. H. Sofia Pinto, A. G´omez-P´erez, J. P. Martins. Some Issues on Ontology Integration. In Proc. of IJCAI99's Workshop on Ontologies and Problem Solving Methods: Lessons Learned and Future Trends, 1999.
16. Karray M H, Morello-Chebel B, Zerhouni N, "towards a maintenance semantic architecture ", Proceedings of The Fourth World Congress on Engineering Asset Management (WCEAM 2009) Athens.
17. www.smac.fr

18. Matsokis, A., Karray, M.H., Morello-Chebel, B. and Kiritsis, D. (2010a). An Ontology-based Model for providing Semantic Maintenance. Proceedings of the 1st IFAC workshop on Advanced Maintenance Engineering, Services and Technology (A-MEST'10); 1-2 July 2010, Lisbon, Portugal.
19. Yoshinobu Kitamura and Riichiro MizoguchiTwelfth, "Functional Ontology for Functional Understanding", International Workshop on Qualitative Reasoning (QR-98), Cape Cod, USA, May 26-29, AAAI Press, pp.77-87, 1998
20. Yoshinobu Kitamura and Riichiro Mizoguchi, "An Ontological Analysis of Fault Process and Category of Faults", Proc. of Tenth International Workshop on Principles of Diagnosis (DX-99), pp.118-128, June 8-11 1999 DX 99 Loch Awe, Scotland 118
21. Ikeda, M., Seta, K., and Mizoguchi, R. (1997) Task Ontology Makes It Easier To Use AuthoringTools. Proc. of IJCAI-97, Nagoya, Japan, 342-347.
22. http://www.isi.edu/isd/LOOM/PowerLoom/
23. Brank, Janez; Marko Grobelnik; Dunja Mladenić. 2005b. A survey of ontology evaluation techniques. SiKDD05.
24. Tatir, Samir; Arpinar, I. Budak. Ontology Evaluation and Ranking using OntoQA. Int. Conf. on Semantic Computing(ICSC), 2007

Résumé

Le progrès rapide des technologies de l'information et de la communication a donné lieu à une variété de systèmes d'aide à la maintenance et des outils couvrant l'ensemble des sous-domaines de la maintenance. Les pluparts de ces systèmes sont basés sur différents modèles qui sont parfois redondantes ou incohérentes et toujours hétérogène. Afin de créer un langage commun et partagé, nous avons développé ontologie de domaine pour la maintenance industrielle en adoptant l'approche METHONTOLOGY pour gérer le développement du cycle de vie de cette ontologie que nous avons appelé IMAMO (). L'ontologie développée peut être exploité pour assurer l'interopérabilité sémantique et de générer de nouvelles connaissances qui supporte la prise de décision dans le processus de maintenance

Une étude comparative diachronique du féminisme

Stéphanie Léon*, Mathieu Roche**

* LIRMM, Université Montpellier 2, France
Université de Provence, 29 avenue Robert Schuman, 13621 Cedex 1
stephanie.leon@lirmm.fr
** LIRMM, Université Montpellier 2, France
mroche@lirmm.fr

Résumé. Cet article présente une approche lexicale d'analyse comparative diachronique entre deux corpus traitant du féminisme, sur deux périodes différentes. L'analyse lexicale s'appuie sur la collecte des « mondes lexicaux » (unités lexicales simples et complexes qui sont significativement fréquentes) liés aux deux corpus et sur une analyse comparative de ces mondes lexicaux. Les résultats montrent que les unités lexicales simples sont très proches entre les deux corpus qui traitent de la même thématique, tandis que les unités lexicales complexes sont significativement différentes, car plus spécialisées à une sous-thématique et à une période.

1. Introduction

Le Centre d'Etudes Alexandrines[1] a entrepris un vaste travail de numérisation de la presse francophone d'Egypte, sur une période de deux cents ans, depuis l'importation de la première presse par Bonaparte en juillet 1798. L'objectif est de mettre à la disposition de la communauté des chercheurs les journaux, revues, périodiques francophones publiés sur le sol égyptien, comme par exemple le Courrier et la Décade égyptienne publiés par Bonaparte entre 1798 et 1801 ou encore la Réforme illustrée des années 1950. L'aspect éphémère de la masse de documents contenue dans ces publications entraînait le risque qu'ils soient négligés voire oubliés par les historiens. Pourtant ces documents contiennent des informations au jour le jour sur l'histoire de l'Égypte sous toutes ses facettes.

Déjà plus d'une dizaine de milliers de pages sont disponibles, non pas en mode image, mais en mode de texte intégral (au format PDF). Suite à ce projet, un travail d'analyse lexicale automatisée s'est mis en place, avec pour objectif d'extraire les « mondes lexicaux » de ces données, selon leur thématique, leur période, etc., en vue de permettre des recherches lexicales et des analyses comparatives. L'objectif de nos travaux est d'extraire les principales caractéristiques lexicales de ces revues, en proposant une méthodologie réapplicable par la suite sur un grand nombre d'autres données. Pour cette étude, nous nous limitons à trois revues (deux revues francophones égyptiennes datant respectivement des années 1930 et 1945 et l'autre, contemporaine et française) et à une thématique, le féminisme. Nous analysons ce thème d'une part dans une perspective comparative diachronique, et d'autre part, par analyse thématique contrastive, entre le féminisme et la critique littéraire. Nous faisons l'hypothèse qu'une analyse comparative lexicale permet de faire émerger les similarités thématiques entre deux revues, mais aussi les spécificités lexicales liées à une sous-thématique ou à une période donnée.

2 Objectifs

L'intérêt de notre projet est de permettre au chercheur (qu'il soit historien, linguiste, etc.) de pouvoir regrouper automatiquement plusieurs revues en fonction de critères précis, comme par exemple en fonction d'un événement (constituer une sélection d'articles transversaux sur un même événement de l'histoire du pays ou de l'histoire de la presse du pays), ou en fonction d'une thématique (la censure, les campagnes de presse, l'insulte et la diffamation…). Il s'agit également d'avoir accès au monde lexical d'une revue afin de connaître ses unités lexicales significatives.

Les aspects lexicaux que nous faisons émerger sont de deux ordres. Nous distinguons deux phénomènes de représentation du contexte d'un mot ou d'une combinaison lexicale, regroupés sous la notion de « monde lexical». D'une part, nous faisons émerger les « mondes lexicaux » des différents corpus étudiés, par une extraction

[1] http://www.cealex.org/

d'Unités Lexicales Simples, répertoriées par catégorie grammaticale « pertinente ». Ces co-occurrences constituent l'entourage lexical, sans prendre en compte les relations de dépendance syntaxique entre les unités lexicales. Nous faisons l'hypothèse que ces mondes lexicaux, représentatifs d'une thématique donnée et du vocabulaire pertinent, permettent d'extraire des régularités entre plusieurs corpus, mais aussi de faire émerger des différences, qu'elles soient sémantiques ou diachroniques.

Dans cette étude, nous avons choisi d'étudier la même thématique, mais à une période différente. Notre comparaison des mondes lexicaux se veut donc diachronique. D'autre part, nous nous intéressons aux patrons morpho-syntaxiques (relations de dépendance syntaxique) les plus récurrents, pour chaque corpus. Dans cet article, nous parlerons d'Unité Lexicale Complexe afin de désigner ces associations lexicales. La section suivante décrit la méthode d'extraction des mondes lexicaux formés des Unités Lexicales Simples et Complexes. La section 3 présente une analyse des mondes lexicaux obtenus. Enfin, quelques perspectives sont présentées en section 4.

3 Extraction des Mondes Lexicaux

3.1 Processus global d'extraction des Mondes Lexicaux

Le processus global proposé est composé de quatre étapes successives. La première étape consiste à acquérir le corpus. Cette phase d'acquisition est détaillée dans la section 4.1 de cet article. La deuxième étape effectue une tâche de normalisation du corpus par un processus de « nettoyage » et d' « uniformisation » des données. L'étape suivante (troisième étape) consiste à étiqueter grammaticalement le corpus avec l'étiqueteur de Brill (Brill, 1994) et/ou le TreeTagger (Schmid, 1994). A partir du corpus étiqueté, la dernière étape extrait les unités lexicales les plus fréquentes. Notons qu'une phase d'analyse des unités lexicales obtenues peut également être ajoutée dans le processus. Cette phase est décrite dans la section 4 de cet article.

3.2 Extraction des Unités Lexicales Simples (ULS)

Le Nous parlons de « monde lexical » afin de désigner les mots-clés les plus fréquents au sein d'une collection de textes. Les mondes lexicaux ont fait l'objet de différentes études, sous des appellations et des applications diverses. La terminologie est un peu floue afin de désigner ce même phénomène. Certains parlent d'isotopie sémantique (Greimas, 1986), de mots-clés thématiques (Rossignol et Sébillot, 2003), de vecteurs conceptuels (Schwab et al., 2004), de signatures thématiques (Lin et Hovy, 2000), ou encore de cartographie lexicale (Véronis, 2003). Les applications relatives à ces notions sont de divers ordre : la Traduction Automatique (Tanguy, 1997 ; 1999) ; la désambigüisation lexicale (Pichon et Sébillot (1999) ; Rossignol et Sébillot (2003)) ; l'enrichissement d'ontologies (Agirre et al., 2000, Agirre et Lopez, 2004) ; la représentation sémantique (Schwab et al., 2004) ; le ré-

sumé automatique (Riloff, 1996, Riloff et Lorenzen, 1999, Hovy et Lin, 1999, Lin et Hovy, 2000).

Notre objectif est d'extraire les mondes lexicaux de chaque corpus (Léon, 2006), en faisant émerger des similitudes et des différences lexicales. En ce qui concerne les catégories grammaticales étudiées, nous faisons l'hypothèse que les noms, les adjectifs et les Entités Nommées (unités simples) sont lexicalement les plus pertinents.

Ces trois catégories ont été obtenues grâce à un étiquetage morpho-syntaxique proposé par le logiciel TreeTagger[2]. Les résultats obtenus en sortie se présentent sous la forme de trois colonnes, avec un terme par ligne et les informations de lemme, de forme et de catégorie grammaticale sur chaque colonne. Le tableau 1 présente un exemple de résultats obtenus avec TreeTagger.

Des	PRP :det	du
dizaines	NOM	Dizaine
de	PRP	De
messages	NOM	message
des	PRP :det	du
comités	NOM	comité
partout	ADV	partout
en	PRP	En
France	NAM	France
nous	PRO :PER	nous
sont	VER:aux:pres	être
parvenus	VER :pper	parvenir

TAB. 1 – *Exemple de résultat obtenu par TreeTagger*

Pour chaque catégorie grammaticale pertinente, nous avons extrait les n mots les plus fréquents de chaque corpus, ce qui a fait émerger le monde lexical classé par catégorie grammaticale.

Le tableau 2 présente les vingt noms les plus fréquents du monde lexical pour le corpus Clara. Ce dernier qui traite du féminisme est décrit de manière précise dans la

[2] http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/

section 4. Cette table montre que les mondes lexicaux sont souvent liés à une thématique donnée.

femme	loi
an	année
droit	violence
vie	jour
homme	temps
pays	personne
monde	société
enfant	association
magazine	famille
fille	question

TAB. 2 – *Monde lexical des noms les plus fréquents du corpus « Clara »*

3.3 Extraction des Unités Lexicales Complexes (ULC)

Nous proposons par la suite d'extraire la terminologie, c'est-à-dire les cooccurrences lexicales entre deux lexèmes liés syntaxiquement et dont la fréquence est significative au sein d'un corpus. Les termes extraits appelés dans la suite des Unités Lexicales Complexes forment un monde lexical spécifique.

De multiples approches de recherche terminologique ont été développées afin d'extraire les termes pertinents à partir d'un corpus. Nous ne traiterons pas ici les approches d'aide à la structuration et au regroupement conceptuel des termes qui sont détaillés dans les travaux de (Aussenac-Gilles et Bourigault, 2003). Les méthodes d'extraction de la terminologie sont fondées sur des méthodes statistiques et/ou syntaxiques. Le système TERMINO de (David et Plante, 1990) est un outil précurseur qui s'appuie sur une analyse syntaxique afin d'extraire les termes nominaux. Cet outil effectue une analyse morphologique à base de règles, suivie de l'analyse des collocations nominales fondée sur une grammaire. Les travaux de (Smadja, 1993) (approche XTRACT) s'appuient sur une méthode statistique. XTRACT extrait, dans un premier temps, les collocations binaires situées dans une fenêtre de dix mots. Les collocations binaires sélectionnées sont celles qui dépassent d'une manière statistiquement significative la fréquence due au hasard. L'étape suivante consiste à extraire les collocations plus générales (collocations de plus de deux mots) contenant les collocations binaires trouvées à la précédente étape. ACABIT de (Daille, 1994) effectue une analyse linguistique afin de transformer les collocations nominales en termes binaires. Ces derniers sont ensuite triés selon des mesures sta-

tistiques. Le système EXIT (Roche, 2004) consiste à extraire les termes complexes de manière itérative en utilisant des critères statistiques (mesures statistiques) et syntaxiques (patrons syntaxiques). Contrairement à ACABIT et EXIT qui sont essentiellement fondés sur des méthodes statistiques, LEXTER (Bourigault, 1993) et SYNTEX (Bourigault et Fabre, 2000) s'appuient en grande partie sur une analyse syntaxique approfondie afin d'extraire la terminologie du domaine. La méthode consiste à extraire les syntagmes nominaux maximaux. Ces syntagmes sont alors décomposés en termes de « têtes » et d' « expansions » à l'aide de règles grammaticales. Les termes sont alors proposés sous forme de réseau organisé en fonction de critères syntaxiques.

Dans notre étude, nous allons appliquer le système EXIT qui a une méthode mixte (syntaxique et statistique) afin d'extraire la terminologie nominale de base[3]. Nous nous intéressons à trois patrons morpho-syntaxiques : Nom-Adjectif, Adjectif-Nom et Nom-Préposition-Nom. Les corpus ont été étiquetés avec l'étiqueteur Brill afin d'extraire les unités respectant ces patrons. Le tableau 3 présente un exemple d'Unités Lexicales Complexes obtenues pour le patron Nom-Adjectif à partir du corpus Clara. Les Unités Lexicales Complexes obtenues sont analysées dans la section suivante.

mutilations génitales	communauté internationale
gynécologie médicale	volonté politique
temps partiel	concept rétrograde
junte militaire	scène slam
commission paritaire	planning familial
parlement européen	comités locaux
état civil	opinion publique
droits humains	condition féminine
acteurs sociaux	justice sociale
prisonniers politiques	journée internationale

TAB. 3 – *Unités Lexicales Complexes (Nom-Adjectif) du corpus « Clara ».*

4 Comparaison des mondes lexicaux des différentes périodes

Dans cette section, nous avons recours à une comparaison entre les mondes lexicaux d'Unités Simples et Complexes des deux périodes. Nous avons extrait les n

[3] Notons que nous n'appliquerons pas le processus itératif d'EXIT.

premières unités, pour chaque catégorie et chaque corpus et nous avons évalué le pourcentage d'unités communes. Nos résultats sont analysés en deux temps. D'une part, nous avons comparé automatiquement le nombre d'unités communes. D'autre part, nous avons analysé manuellement la pertinence des unités non communes.

4.1 Description des corpus

Les corpus étudiés concernent deux revues traitant de la même thématique, le féminisme, sur deux époques distinctes. La première revue, « l'Egyptienne », est une revue mensuelle francophone diffusée en Egypte, datant des années 1930, traitant de sujets autour du féminisme tels que la politique, la sociologie, l'art, etc. Nous avons réuni 11 numéros de cette revue, qui nous ont été accessibles grâce au Centre d'Etudes Alexandrines. Ces numéros couvrent les dates de septembre 1925 jusqu'à mars 1930. Ils représentent environ 250 000 mots.

Cette revue a été numérisée par le Centre d'Etudes Alexandrines, à l'aide d'outils efficaces pour la Reconnaissance Optique de Caractères (ROC) (par exemple un appareil de prise de vues 'Phase One' qui permet de scanner de grands formats avec une forte précision pour ensuite pouvoir procéder à une reconnaissance de caractères avec une grande fiabilité). La reconnaissance optique de caractères permet ensuite de traduire des images de textes imprimés ou dactylographiés en fichiers de texte. Les fichiers obtenus par le Centre d'Etudes Alexandrines sont au format PDF. Nous les avons convertis en mode texte grâce à un logiciel de conversion de fichiers PDF en TXT[4]. Cette conversion a posé quelques difficultés liées à l'exploitation automatique de corpus « océrisés » (altération de certaines chaînes de caractères) et nous avons procédé à une phase manuelle de nettoyage.

La deuxième revue, « Clara », est une revue sur le féminisme dont les archives sont en ligne[5]. Nous avons réuni les archives disponibles, traitant de diverses thématiques telles que le racisme, l'Europe, la violence, etc. Les documents étant plus courts que pour la revue « l'Egyptienne », nous avons réuni 84 archives, qui se présentent sous la forme de dossiers traitant d'un sujet d'actualité donné. Les périodes s'étendent de septembre 2006 jusqu'au mois de janvier 2010. Le nombre de mots est d'environ 100 000.

Nous utilisons un autre corpus de test, qui ne traite pas de la même thématique (sans en être complètement éloigné), afin de pouvoir établir des comparaisons de résultats. Il s'agit de la revue francophone égyptienne « Valeurs », cahier trimestriel de critique et de littérature, datant d'avril 1945 à janvier 1947. Le nombre de mots est d'environ 400 000.

---

[4] http://www.simpopdf.com/pdf-to-text.html
[5] http://clara-magazine.fr/

4.2 Résultats

4.2.1 Analyse quantitative des unités communes

Le tableau 4 montre une différence entre les mondes lexicaux des unités complexes et ceux des unités simples : les mondes lexicaux des unités complexes sont relativement éloignés contrairement à ceux des unités simples. Le fait que les unités simples, et plus particulièrement les noms et adjectifs, soient très proches confirme que nous étudions une même thématique (par exemple, les unités simples femme, droit, travail, enfant, famille, significativement pertinentes pour notre thématique sont communes aux deux corpus).

Cependant, les Entités Nommées (EN) qui sont souvent liées à une époque restent naturellement assez spécifiques. Il en va de même pour l'analyse contrastive des Unités Lexicales Complexes, qui met en exergue des préoccupations et des spécificités d'une époque. Ceci explique donc les pourcentages d'unités complexes communes très faibles pour les unités complexes (cf tableau 4).

Ces unités seront évaluées dans la section suivante qui propose une analyse contrastive à partir des unités non communes entre les deux corpus (« Clara » et « l'Egyptienne »). Ceci nous permettra de vérifier si de telles unités sont liées au domaine du féminisme en mettant en relief les unités propres à une époque.

n	**Unités lexicales complexes (ULC)**			**Unités lexicales simples**		
	Nom-Adj	**Adj-Nom**	**Nom-prép-Nom**	**Adj**	**Nom**	**EN**
10	0%	50%	0%	70%	70%	40%
50	8%	24%	0%	64%	42%	24%
100	5%	15%	2%	50%	41%	31%
200	2.5%	9.5%	1,5%	50%	43%	42%

TAB. 4 – *Pourcentage d'unités communes parmi les n premières unités les plus fréquentes extraites (« Clara » et « l'Egyptienne »).*

Le tableau 5 montre le pourcentage d'unités communes entre le corpus « Clara » et l'autre corpus que nous étudions, « Valeurs ». Nous remarquons que les noms, principalement, sont nettement moins communs entre les deux corpus, essentiellement pour les 50 premiers noms. Les adjectifs se maintiennent à environ la moitié d'adjectifs communs. La proportion d'unités communes s'explique par le fait que la thématique du féminisme peut aborder certains aspects relatifs à la littérature. Toutefois, la différence entre la proportion d'unités simples communes entre le corpus «

Clara » et « l'Egyptienne » d'une part, et « Clara » et « Valeurs » d'autre part est significative, ce qui confirme notre hypothèse de similarité lexicale pour une thématique donnée.

Le tableau 5 montre également que les unités complexes Nom-Adjectif à partir des corpus « Clara » et « Valeurs » sont très différents, ce qui confirme notre hypothèse selon laquelle les unités complexes sont représentative d'une époque et d'une thématique. Ici, la comparaison concerne bien deux périodes et deux thématiques différentes.

n	**Unités lexicales complexes (ULC)**	**Unités lexicales simples (ULS)**	
	Nom-Adj	**Adj**	**Nom**
10	0%	50%	30%
50	0%	52%	22%
100	1%	49%	34%
200	1%	53%	36%

TAB. 5 – *Pourcentage d'unités communes parmi les n premières unités les plus fréquentes extraites, entre deux corpus ne traitant pas du même thème (« Clara » et « Valeurs »).*

Notons que les résultats sur les unités lexicales complexes Adjectif-Nom et Nom-Préposition-Nom n'ont pas été reportés dans le tableau (tableau 4). En effet les résultats obtenus sont à nuancer car le nombre d'occurrences de toutes les unités est moindre (égal à 1). Dans ce cas, le classement par nombre d'occurrences n'est pas un critère adapté car l'ordre proposé devient en fait aléatoire. Si nous considérons les 200 premières unités Adjectif-Nom et Nom-Préposition-Nom, nous n'obtenons aucun terme commun. Si nous sommes en condition idéale (tous les termes communs extraits avec notre système placés en tête), le pourcentage de termes communs sur les 200 premiers termes des unités Adjectif-Nom et Nom-Préposition-Nom est respectivement de 5,5% (11 termes en commun sur 200) et de 0%. Ces proportions restent très faibles et confirment que les unités lexicales complexes des deux corpus sont très différentes.

Dans la section suivante, nous verrons l'analyse quantitative de ces résultats.

4.2.2 Analyse quantitative des unités communes

Le tableau 6 montre que le corpus « Clara » possède une proportion d'unités pertinentes spécifiques liées au féminisme plus importante comparativement au corpus « l'Egyptienne » (49% et 62% pour les Noms et Noms-Adjectifs du corpus « Clara »

VS 37% et 40% pour le corpus « l'Egyptienne »). La Figure 6 montre également que les unités complexes non communes sont plus largement liées au féminisme que les unités simples pour chaque corpus. Ces résultats confirment que les unités complexes sont, naturellement, plus spécifiques à une thématique et à une époque que les unités simples, comme par exemple « planning familial », « orientation sexuelle », « harcèlement sexuel », etc. pour le corpus Clara et « mortalité infantile », « fièvre puerpérale », etc. pour le corpus l'Egyptienne.

Pertinence des Unités	**Nom**		**Nom-Adjectif**	
	Clara	**Egyp-tienne**	**Clara**	**Egyp-tienne**
Unités pertinentes (liées au féminisme)	49%	37%	62%	40%
Unités pertinentes mais trop générales (non liées au féminisme)	44%	59%	38%	59%
Unités non pertinentes	7%	3%	0%	1%

TAB. 6 – *Pourcentage des 100 premières unités non communes évaluées manuellement : unités de type Nom (ULS) et Nom-Adjectif (ULC) extraites à partir des deux corpus (« Clara » et « l'Egyptienne »).*

En ce qui concerne l'analyse qualitative des unités non communes entre le corpus « Clara » et le corpus « Valeurs » (tableau 7), nous notons qu'une infime partie des noms du corpus « Valeurs » est liée au féminisme, contrairement aux autres corpus.

Pertinence des Unités	**Nom**	
	Clara	**Valeurs**
Unités pertinentes (liées au féminisme)	56%	6%
Unités pertinentes mais trop générales (non liées au féminisme)	44%	94%
Unités non pertinentes	0%	0%

TAB. 7 – *Pourcentage des 100 premières unités non communes évaluées manuellement : unités de type Nom (ULS) extraites à partir des deux corpus (« Clara » et « Valeurs »).*

6 Conclusion

Nous avons présenté une analyse lexicale diachronique entre deux corpus traitant de la même thématique, sur une période différente. L'analyse lexicale s'est appuyée sur l'émergence et la comparaison des mondes lexicaux des deux corpus. Ces mondes lexicaux, unités lexicales les plus fréquentes, mettent en valeur la thématique et l'univers lexical d'un corpus. La comparaison de ces mondes lexicaux permet d'une part de confirmer la similarité thématique entre deux corpus et d'autre part de mettre en valeur les spécificités de chaque corpus, qu'il s'agisse de sous-thématiques spécialisées ou de divergences diachroniques. Ce type d'étude pourrait être utile dans d'autres contextes applicatifs, comme par exemple celui de la désambigüisation lexicale ou de la Traduction Automatique.

Dans nos futurs travaux, nous souhaitons améliorer la phase de nettoyage des données afin d'obtenir des corpus moins bruités. Nous souhaitons par ailleurs effectuer un classement des unités fondé sur des mesures statistiques plus adaptées que la fréquence. En effet, cette dernière ne prend pas en compte la répartition des unités dans les différents documents d'un même corpus contrairement à d'autres critères tels que la mesure TF-IDF.

Références

AGIRRE, E., OLATZ, A., HOVY, E., MARTINEZ, D. (2000). Enriching very large ontologies using the WWW. Ontology Construction of the European Conference of AI (ECAI), Berlin, Allemagne.

AGIRRE, E., LOPEZ, O. (2004). Publicly available topic signatures for all wordnet nominal senses. Proceedings of the 4rd International Conference on Language Resources and Evaluation (LREC), Lisbon, Portugal.

AUSSENAC-GILLES N., BOURIGAULT D. (2003), Construction d'ontologies à partir de textes. Actes de TALN, Volume 2, p27–47.

BOURIGAULT D., FABRE C. (2000), Approche linguistique pour l'analyse syntaxique de corpus. Cahiers de Grammaires 25, p131–151.

BOURIGAULT D. (1993), Analyse syntaxique locale pour le repérage de termes complexes dans un texte. TAL, 34(2), p105–118.

BRILL E. (1994), Some advances in transformation-based part of speech tagging. In AAAI, Vol. 1, pp. 722–727.

DAILLE B. (1994), Approche mixte pour l'extraction automatique de terminologie : statistiques lexicales et filtres linguistiques. Ph.D. thesis, Univ. Paris 7.

DAVID S., PLANTE P. (1990), De la nécéssité d'une approche morpho syntaxique dans l'analyse de textes. In Intelligence Artificielle et Sciences Cognitives au Quebec, Volume 3, pp. 140–154.

HOVY, E., LIN C. Y. (1997). Automated Text Summarization in SUMMARIST. Workshop on Intelligent Scalable Text Summarization, Madrid, Espagne.

LEON S. (2006), Acquisition automatique de traductions de termes complexes par comparaison de « mondes lexicaux » sur le Web. Actes de RECITAL, p700-709.

LIN C.-Y., HOVY E. (2000), The Automated Acquisition of Topic Signatures for Text Summarization. Actes de COLING.

PICHON R., SÉBILLOT P. (1999). Différencier les sens des mots à l'aide du thème et du contexte de leurs occurrences : une expérience. Actes de Traitement Automatique des Langues Naturelles (TALN).

RILOFF E., LORENZEN J. (1998). Extraction-Based Text Categorization: Generating Domain-Specific Role Relationships Automatically, Natural Language Information Retrieval, p167-196

ROCHE M. (2004), Intégration de la construction de la terminologie de domaines spécialisés dans un processus global de fouille de textes. Ph.D. thesis, Univ. Paris 11.

ROSSIGNOL M., SEBILLOT P. (2003), Extraction statistique sur corpus de classes de mots-clés thématiques. TAL, 44(3), p217-246.

SMADJA F. (1993), Retrieving collocations from text : Xtract, Computational Linguistics, Vol. 19, pp. 143-177.

SCHWAB S., LAFOURCADE M., PRINCE V. (2004), Hypothèses pour la construction et l'exploitation conjointer d'une base lexicale sémantique basée sur les vecteurs conceptuels. Actes des JADT, Louvain-la-Neuve, Belgique.

TANGUY, L. (1997). Traitement automatique de la langue naturelle et interprétation : contribution à l'élaboration informatique d'un modèle de la sémantique interprétative. Thèse de doctorat, Ecole Nationale Supérieure des Télécommunication de Bretagne.

TANGUY L. (1999). Isotopies sémantiques pour la vérification de traduction. Traitement Automatique des Langues Naturelles (TALN).

SCHMID H. (1994), Probabilistic Part-of-Speech Tagging Using Decision Trees. In Proceedings of the Int. Conf. on New Methods in Language Processing, p44-49.

VERONIS J. (2003), Cartographie lexicale pour la recherche d'information. Actes de TALN, Batz-sur-Mer, France

Remerciements

Un grand merci à Louis-Jean Calvet (Université de Provence), au Centre d'Etudes Alexandrines et particulièrement à Jean-Yves Empereur, pour avoir initié ce projet et mis à notre disposition toutes les ressources numérisées de la presse francophone.

Summary

This paper presents a diachronic comparative analysis between two corpora dealing with the domain of feminism, on two different periods. Lexical analysis is based on the acquistion of "lexical worlds" (i.e. simple and complex lexical units significantly frequent) related with both corpora and on a comparative analysis of these worlds. The results show that the simple lexical units are very similar between both corpora that deal with the same topic, while the complex lexical units are significantly different, because they are more specialized to a sub-topic and a period.